REMARKS

Claims 1-57, 62, 63, 69, 70 and 159-164 were pending in the application. Claims 9, 10, 11, 12, 15, 16, 21, 22, 24, 25, 38 and 62 have been amended. New claims 165-206 have been added. Claims 165-206 were added to make the pending multiply dependent claims (*i.e.*, claims 10-12, 15-29 and 62) singly dependent. Claims 42-57, 63, 69, 70 and 159-164 were withdrawn from consideration as being directed to a non-elected invention and are hereby cancelled without prejudice. Accordingly, upon entry of the present amendment claims 1-41, 62 and 165-206 will remain pending in the instant application.

Support for the amendments to the claims may be found throughout the specification and in the claims, as originally filed. In particular, support for the amendment to claim 9 may be found in claim 8 as originally filed and at, for example, page 29, line 24 through page 30, line 2 of the specification. New claims 165-206 have been added in view of the amendments to the multiply dependent claims. *No new matter has been added.*

Any amendments to and/or cancellation of the claims are not to be construed as an acquiescence to any of the rejections set forth in the instant Office Action, and were done solely to expedite prosecution of the application. Applicants hereby reserve the right to pursue the subject matter of the claims as originally filed in this or a separate application(s).

Interview

Applicants greatly appreciate the Examiner's availabilty to discuss the outstanding Office Action in a telephinic interview held on November 18, 2004.

Information Disclosure Statement

A Supplemental Information Disclosure Statement is being submitted on even date herewith (*via* First Class Mailing).

Specification –Sequence Rules

The Examiner objects to the specification for disclosing nucleotide and/or amino

acid sequences on Figures 1, 2, 9, 10; Table 13 and 14 which fail to comply with the requirements of 37 CFR 1.821 through 1.825. The specification and figure legends have been amended to include sequence identifiers, preceded by "SEQ ID NO:" as required by 37 C.F.R. 1.821(d). No new matter has been added.

The Examiner objects to the disclosure for the incorrect citation of WIPO publication number "W087/02671" on page 42, line 10, and for a double underlined heading on page 108, line 2. Applicants have made the appropriate correction to the WIPO publication number and heading as requested by the Examiner. Accordingly, the foregoing issue has been rendered moot.

Claim Objection

The Examiner has objected to claim 62 for referring to non-elected claims. Claim 62 has been appropriately amended.

Double Patenting

The Examiner has *provisionally* rejected claims 1-41 and 62 under 35 U.S.C. 101 as being drawn to the "same invention" as that of claims 1-41 and 62 of copending Application No. 10/232,030. The Examiner has also *provisionally* rejected claims 1-41 and 62 under 35 U.S.C. 101 as being drawn to the "same invention" as that of claims 1-41 and 62 of copending Application No. 10/388,389. Applicants respectfully traverse. None of claims 1-41 and 62 of copending Application No. 10/232,030 and claims 1-41 and 62 of copending Application No. 10/388,389 have been allowed. Applicant's seek allowance of the enumerated claims in the instant application and intend to cancel any conflicting subject matter from each of Application Nos. 10/232,030 and 10/388,389 as appropriate. Withdrawal of the provisional rejections under 35 U.S.C. 101 is respectfully requested.

The Examiner has also *provisionally* rejected claims 1-41 and 62 under the judicially created doctrine of obviousness-type double patenting as being unpatentable

over claims 1-11 and 23 of copending Application No. 10/703,713. This is a *provisional* obviousness-type double patenting rejection because the conflicting claims have not in fact been patented. Applicants respectfully traverse. None of claims 1-11 and 23 of copending Application No. 10/703,713 have been allowed. Applicant's seek allowance of the enumerated claims in the instant application and, if appropriate, intend to address any obviousness-type double patenting issues in Application No. 10/703,713. Allowance of the pending claims and withdrawal of the provisional obviousness-type double patenting rejection over claims 1-11 and 23 of copending Application No. 10/703 is respectfully requested.

The Examiner has also *provisionally* rejected claims 1-41 and 62 under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 47, 67, 69 and 70 of copending Application No. 09/724,552. As discussed in an interview with the Examiner on November 18, 2004, Application No. 09/724,552 has now issued as U.S. Patent No. 6,750,324 (hereinafter "the '324 patent"). For the purposes of this argument, Applicants will treat the provisional rejection of claims 1-41 and 62 as an actual obviousness-type double patenting rejection since the conflicting claims have, in fact, been patented. Also, for the purposes of this argument, Applicants submit that claims 47, 67, 69 and 70 of previously co-pending Application No. 09/724,552 correspond to issued claims 1, 8, 2 and 3, respectively, of the '324 patent.

Applicants respectfully traverse this rejection and request reconsideration and withdrawal on the grounds that claims 1-41 and 62 are directed to patentably distinct species of humanized immunoglobulins. Claims 1, 8, 2 and 3 of the '324 patent are directed to pharmaceutical compositions and diagnostic kits comprising a chimeric or humanized antibody that specifically binds to an epitope within residues 1-10, 1-6 or 1-4 of Aβ. In contrast, claims 1-41 and 62 are directed to patentably distinct species of humanized immunoglobulins comprising CDRs from a specific mouse antibody, the 3D6 antibody, framework regions from a human acceptor and at least one framework residue substitution from a specified set of residues with the corresponding amino acid residue from the 3D6 light or heavy chain variable region sequence.

Significantly, the differences in subject matter which render pending claims 1-41 and 62 patentably distinct from claims 1, 8, 2 and 3 of the '324 patent include (1) the presence of CDRs from either the light or heavy chain of the 3D6 antibody set forth in SEQ ID NO: 2 (light chain) or SEQ ID NO: 4 (heavy chain); and (2) framework regions from a human acceptor having at least one framework residue substitution from a specified set of residues with the corresponding amino acid residue from the 3D6 light or heavy chain variable region sequence. Additional differences in subject matter which render the claims patentably distinct are enumerated in each of the pending independent and dependent claims. Specifically, independent claims 1, 2, 6 and 7 further specify that the framework residue substitution be a residue that either non-covalently binds antigen directly, is adjacent to or interacts with a CDR, participates in the VL-VH interface (claims 1 and 2) or is capable of affecting light or heavy chain variable region conformation or function as identified by analysis of a three-dimensional model of the 3D6 immunoglobulin light or heavy chain variable region (claims 6 and 7). Independent claims 13 and 14 recite that the framework substitution be at a site selected from residue L1, L2, L36, or L46 of the light chain (claim 13) or H49, H93 or H94 of the heavy chain (Kabat numbering convention) (claim 14). Independent claims 30 and 31 specify that each of the framework substitutions L1, L2, L36, or L46 of the light chain (claim 30) or H49, H93 or H94 of the heavy chain (Kabat numbering convention) (claim 31) be present. Each of dependent claims 3-5, 8-12, 15-29, 32-41 and 62 include one or more of these recitations as well as additional differences in subject matter from claims 1, 8, 2 and 3 of the '324 patent which renders the claims patentably distinct. These differences are set forth in each of the dependent claims, the substance of which is reiterated here in support of Applicants' traversal.

Accordingly, none of the differences in subject matter recited in claims 1-41 and 62, in particular the CDRs from the 3D6 antibody and the framework regions from a human acceptor having a framework substitution from a specified set of framework residues with the corresponding amino acid residue from the 3D6 antibody, is obvious over claims 1, 8, 2 and 3 of the '324 patent. As such, each of the pending claims 1-41

and 62 is patentably distinct over claims 1, 8, 2 and 3 of the '324 patent and it is respectfully requested that the obviousness-type double patenting rejection be reconsidered and withdrawn.

Rejection of Claims 8, 9, 21, 22, 24, 25 and 32 Under 35 U.S.C. 112, Second Paragraph

The Examiner has rejected claims 8, 9, 21, 22, 24, 25 and 32 under 35 U.S.C. 112, second paragraph as allegedly being indefinite over the recitation of the terms "interchain packing residue" (claims 8 and 9), "rare residue" (claim 8), "unusual residue" (claim 9), "rare human framework residue" (claims 21, 22, 24 and 25) and "framework residue" (claim 32). Applicants traverse.

Claim 9, as amended, no longer recites the term unusual thus obviating the rejection in part.

Regarding the term "interchain packing residue", Applicants respectfully submit that the term has an art-recognized meaning that is clear to artisans skilled in the field of antibody structure. The instant specification defines the term "interchain packing residue" in a manner consistent with the art recognized meaning, see *e.g.*, page 28, lines 22-26 where an "interchain packing residue" is defined as a residue at the interface between VL and VH. The specification cites two seminal references which describe the key role of "interchain packing residues" in antibody structure and function. The references cited are provided herewith for the Examiner's convenience as APPENDICES A and B. (See C. Chothia, J. Novotny, R. Bruccoleri and M. Karplus, *Domain association in immunoglobulin molecules: The packing of variable domains* J. Mol. Biol. 186:651-663 (1985); and J Novotný and E Haber. *Structural invariants of antigen binding: comparison of immunoglobulin VL-VH and VL-VL domain dimers*. Proc Natl Acad Sci U S A. 1985 July; 82(14): 4592–4596. Moreover, the instant specification exemplifies the identification of interchain packing residues in the 3D6 antibody (see *e.g.*, Figure 1, and Tables 12 and 13). In view of the foregoing, it is Applicants' position that the meaning of the term "interchain packing residue" is definite and respectfully requests reconsideration of the rejection under 35 U.S.C. 112, second paragraph.

Regarding the term "rare", Applicants respectfully submit that the term has an art-recognized meaning that is clear to artisans skilled in the field of antibody structure. The instant specification defines the term "rare" in a manner consistent with the art recognized meaning and teaches the artisan how to identify such residues, see e.g., page 29, line 10 through page 30, line 16. Rare residues are identified by comparison of the sequence of an antibody of interest with the sequences of known antibodies of a similar type (*e.g.*, species, subtype). An extensive list of known sequences is found in Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C. *Sequences of Proteins of Immunological Interest*, 5th ed. 1, U.S. Department of Health and Human Services, National Institute of Health, Bethesda, MD (NIH Publication No. 91-3242) 1991, a standard reference relied on by artisans skilled in the art of antibody structure. In view of the foregoing, it is Applicants' position that the meaning of the term "rare" is definite and respectfully requests reconsideration of the rejection under 35 U.S.C. 112, second paragraph.

Regarding the term "framework residue", Applicants respectfully submit that the term has an art-recognized meaning that is clear to artisans skilled in the field of antibody structure, namely, a residue within an antibody framework region. The instant specification teaches the organization of antibody variable regions as complementarity determining regions ("CDRs") (also referred to in the art as "hypervariable regions" interspersed with the framework regions ("FRs") at, for example page 19, line 20 through page 20, line 6. This teaching is consistent with the art-recognized description of antibody structural organization. A textbook chapter, *Chapter 3, Immunoglobulins: Structure and Function* from *Fundamental Immunology, 4th Edition*, Lippincott-Raven publishers, is provided for the Examiner's convenience (APPENDIX C). At pages 41-44, the structural organization of antibody variable regions is described with specific description of the boundaries of the CDRs and FRs set forth in Table 1. The definition of CDRs in the variable light (VL) chain was originally described in the seminal reference, Wu and Kabat. *An analysis of the sequences of the variable regions of Bence-Jones proteins and myeloma light chains and their implications for antibody complementarity*. J Exp Med 1970;132:211-250 (copy provided as APPENDIX D, see, in particular, page

237) and is discussed at page 41 of the textbook chapter. Kabat's *Sequences of Proteins of Immunological Interest* or (another seminal reference in the field) can likewise be relied on to identify CDRs (*i.e.*, hypervariable regions) and, hence, framework regions (and framework residues). The identification of framework regions and residues within the 3D6 antibody is exemplified in the instant specification (see *e.g.*, Table 13) in a manner consistent with the art-recognized meaning of the term "framework". In view of the foregoing, it is Applicants' position that the meaning of the term "framework residue" is definite and respectfully requests reconsideration of the rejection under 35 U.S.C. 112, second paragraph.

CONCLUSION

If a telephone conversation with Applicants' Attorney would expedite the prosecution of the above-identified application, the examiner is urged to call the undersigned at (617) 227-7400.

Respectfully submitted,

LAHIVE & COCKFIELD, LLP

Debra J. Milasincic, Esq.
Reg. No. 46,931
Attorney for Applicants

28 State Street
Boston, MA 02109
Tel. (617) 227-7400
Dated: November 26, 2004

Domain Association in Immunoglobulin Molecules

The Packing of Variable Domains

Cyrus Chothia

MRC Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH

and

Christopher Ingold Laboratories
Department of Chemistry, University College London
20 Gordon Street, London WC1H 0AJ, England

Jiři Novotný, Robert Bruccoleri

Molecular & Cellular Laboratory
Massachusetts General Hospital

and

Harvard Medical School, Boston MA 02114, U.S.A.

and

Martin Karplus

Department of Chemistry
Harvard University, Cambridge MA 02138, U.S.A.

We have analyzed the structure of the interface between VL and VH domains in three immunoglobulin fragments: Fab KOL, Fab NEW and Fab MCPC 603. About 1800 Å² of protein surface is buried between the domains. Approximately three quarters of this interface is formed by the packing of the VL and VH β-sheets in the conserved "framework" and one quarter from contacts between the hypervariable regions. The β-sheets that form the interface have edge strands that are strongly twisted (coiled) by β-bulges. As a result, the edge strands fold back over their own β-sheet at two diagonally opposite corners. When the VL and VH domains pack together, residues from these edge strands form the central part of the interface and give what we call a three-layer packing; i.e. there is a third layer composed of side-chains inserted between the two backbone side-chain layers that are usually in contact. This three-layer packing is different from previously described β-sheet packings. The 12 residues that form the central part of the three observed VL–VH packings are absolutely or very strongly conserved in all immunoglobulin sequences. This strongly suggests that the structure described here is a general model for the association of VL and VH domains and that the three-layer packing plays a central role in forming the antibody combining site.

1. Introduction

Immunoglobulins are the best-studied examples of a large and ancient family of proteins, which also includes β-microglobulins. Thy-1 antigens, major (i.e. class I) and minor (i.e. class II) histocompatibility antigens and cell surface receptors. Functionally, all these structures are involved in cell recognition processes (Jensenius & Williams, 1982), either actively as vehicles endowed with

recognition specificity (antigen-combing antibodies) or passively as surface structures that are being recognized (histocompatibility antigens). Only the immunoglobulin tertiary structures are known to date (Schiffer et al., 1983; Epp et al., 1974; Saul et al., 1978; Segal et al., 1974; Marquart et al., 1980; Deisenhofer, 1981; Phizackerley et al., 1979). However, the homology among primary structures of immunoglobulin, β-microglobulin, Thy-1 antigen, some of the histocompatibility antigen domains, T-cell receptor β chain and the transepithelial "secretory component" has been interpreted as evidence for a common fold (Cunningham et al., 1973; Orr et al., 1979; Feinstein, 1979; Cohen et al., 1980, 1981a; Novotný & Auffray, 1984; Yangai et al., 1984; Hedrick et al., 1984; Mostov et al., 1984).

A typical antibody molecule (IgG1) consists of two pairs of light chains (M_r 25,000) and two pairs of heavy chains (M_r 50,000), each of the chains being composed of domains made up of approximately 100 amino acid residues. The domains are autonomous folding units; it has been demonstrated experimentally (Hochman et al., 1973; Goto & Hamaguchi, 1982) that a polypeptide chain segment corresponding to a single domain can be refolded independently of the rest of the polypeptide chain. All the immunoglobulin domains are formed by two β-sheets packed face-to-face and covalently connected together by a disulfide bridge. The topology of the N-terminal, variable domains in both the light and heavy chains differs from that of the C-proximal constant domains. While the two variable domain sheets consist of five and four strands, respectively, the constant domain sheets are three- and four-stranded (Fig. 1). The four-stranded β-sheets of the two domain types are homologous; the five- or four- stranded β-sheet of the variable domains derives from the three-strand sheet of the constant domains by the addition, at one side, of a two-stranded β-hairpin or a single β-strand, respectively.

In a complete immunoglobulin molecule, domains that correspond to different polypeptide chains associate to form domain dimers VL–VH, CL–CH1 and CH3–CH3. Edmundson et al. (1975) were the first to note the phenomenon of rotational allomerism between the variable and constant domain dimers, that is, whereas the C–C dimers interact via a close packing of their four-strand sheets, the V–V dimers pack "inside out", with the five-stranded sheets oriented face-to-face. The reversal of domain–domain interaction is reflected in the amino acid sequence homology between, and among, the constant and variable domains (Novotný & Franěk, 1975; Beale & Feinstein, 1976; Novotný et al., 1977).

Different antibody molecules in the same organism bind different antigenic structures. The variation in specificity is produced by several mechanisms: mutations, deletions and insertions in the binding regions of the VL and VH domains; and the association of different light and heavy chains. Aspects of the second mechanism are analyzed in

this paper. In particular, the nature of the interface between VL and VH domains is examined by comparing the Fab fragments of KOL, NEW and MCPC 603 myeloma proteins whose X-ray structures are known. The relative contributions to the buried surface between the domains from the conserved framework residues and the hypervariable regions are determined. Attention is focused on the unique packing of the interfaces and the reasons for this packing are examined.

2. Materials and Methods

(a) Fab fragment co-ordinates

Cartesian co-ordinates for Fab fragments KOL, NEW and MCPC 603 were obtained from the Brookhaven Data Bank (Bernstein et al., 1977). Table 1 lists the domain classification, the nominal resolutions and the crystallographic residuals (R factors) for the 3 Fab fragments. To facilitate comparisons of the 3 structures, their residue numbering was changed from that used in the original descriptions to that used by Kabat et al. (1983). Thus, in this paper residues that are structurally homologous have the same sequence number.

To obtain consistent sets of atomic co-ordinates, the original co-ordinates were dissected into individual VL–

(a)

(b)
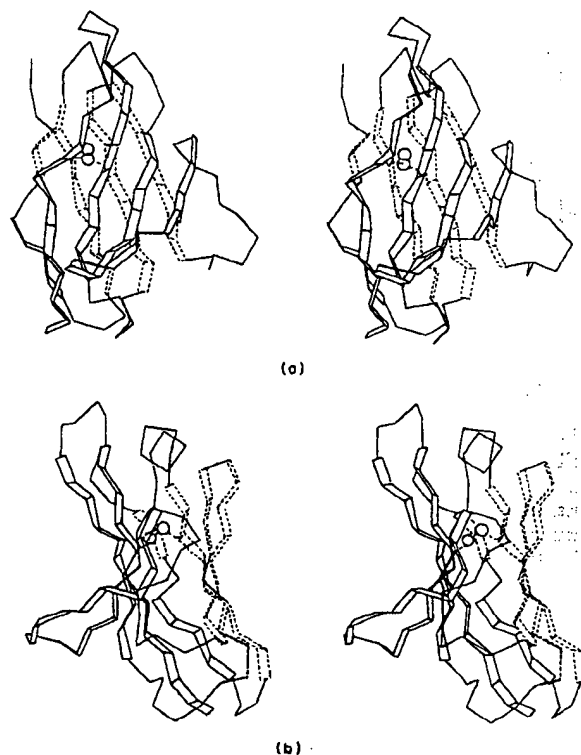
Figure 1. The β-sheets in typical immunoglobulin domains. Vertices represent the position of Cα atoms: those in β-sheets are linked by ribbons; and those between strands by lines. (a) The VL domain of KOL: the β-sheet involved in VL–VH contacts is closer to the viewer (unbroken line). (b) The same VL domain rotated by approximately 90°. Note that the interface-forming β-sheet is strongly twisted at diagonally opposite corners (drawing by A. M. Lesk).

Table 1
Summary of X-ray crystallographic data

Protein	L and H chain types	X-ray data		Minimized		Reference
		Resolution (Å)	R factor (%)	Energy (kJ)	r.m.s. shift (Å)	
Fab KOL human	λI, γIII	1·9	26	−3010	—	Marquart *et al.* (1980)
Fab NEW human	λI, γII	2·0	19	−2592	0·21	Saul *et al.* (1978)
Fab MCPC 603 mouse	κ, γI	2·7	24	−3703	0·26	Segal *et al.* (1974)

The energy given for Fab KOL is that of the unminimized crystallographic data.

VH domain dimers. The structures were subjected to 100 cycles of constrained energy minimization with the program CHARMM version 16 using the adopted-basis Newton-Raphson procedure (Brooks *et al.*, 1983) with constraints of 41·8 kJ (10 kcal) present on all the atoms (Bruccoleri & Karplus, unpublished results). Typically, the constrained minimization converged from original positive values of potential energy to values of about −2·1 kJ/atom (−0·50 kcal/atom) with an average root-mean-square co-ordinate different from the original X-ray structure of 0·3 Å (see Table 1). The results indicate that the crystallographic structures were satisfactory and that acceptable values of potential energy can be achieved by small adjustments of the co-ordinates. Thus, both energy minized structures and the crystallographic co-ordinates were used in the present study; essentially identical results were obtained from the 2 types of co-ordinates sets.

(b) *Computation of solvent-accessible surfaces and contact areas*

Solvent-accessible surfaces (Lee & Richards, 1971) were computed with programs written by A. M. Lesk using the method of Shrake & Rupley (1973) and by T. Richmond using the methods of Lee & Richards (1971) and Richmond & Richards (1978). The latter program was obtained from Yale University. The water probe radius used was 1·4 Å and the section interval along the Z axis was 0·05 Å; the atom van der Waals' radii used were 2 Å for all the (extended) tetrahedral carbon atoms, 1·85 Å for all the planar (*sp2* hybridized) carbons, 1·4 Å and 1·6 Å for carbonyl and hydroxyl oxygens, respectively, 1·5 Å for a carbonyl OH group, 2·0 Å for all the (extended) tetrahedral nitrogen atoms, 1·5 Å, 1·7 Å and 1·8 Å for *sp2*-hybridized nitrogen atoms carrying no hydrogen, 1 and 2 hydrogen atoms, respectively, 2·0 Å for a sulfhydryl group and 1·85 Å for a divalent sulfur atom with no hydrogens.

(c) *β-Strands and β-sheets*

Protein structures were analyzed using the CHARMM program (Brooks *et al.*, 1983) in the so-called explicit hydrogen atom representation: aliphatic hydrogens were combined together with their heavy atoms into "extended atoms" whereas hydrogens bound to polar atoms and possibly involved in hydrogen bonds were explicitly present. The β-strands and β-sheets were defined by their inter-strand backbone (C = O . . . H-N) hydrogen-

bonding pattern. A hydrogen bond list was generated in CHARMM for all the polypeptide chain segments under consideration and amino acids with hydrogen bonds of nearly optimal geometry (energy of −4·18 kJ/bond or less) were taken to be parts of the β-sheets (cf. Fig. 3 of Novotný *et al.*, 1983). This method of defining β-strand boundaries gives results essentially identical to those obtained by visual inspection of crystallographic models, although it tends to be somewhat more restrictive (the 2 methods sometimes differ in inclusion of the N- or C-terminal β-strand residues). Ambiguities arise in cases of edge β-strands that start and end with irregular conformations (β-bulges); such cases are discussed in more detail below.

(d) *β-Strand conformation*

In a typical extended polypeptide chain segment, the dihedral angle between the 2 consecutive side-chains is not 180° as in the ideal β-sheet (Pauling *et al.*, 1951) but closer to −160°; that is, the β-strands are twisted (Chothia, 1973). The out-of-planarity angle (180° − 160°) = 20° can be obtained explicitly from the values of the principal backbone torsion angles φ, ψ and ω (see, e.g. Chou *et al.*, 1982). We define the local backbone twist for 2 consecutive residues as:

$$9 = \left(-\frac{\tau}{|\tau|} \right)(180 - |\tau|),$$

where τ is the torsion angle Cβ–Cα–C'α–C'β and $|\tau|$ denotes its magnitude. When glycine residues that lack Cβ atoms are encountered, the torsion angle 9 is measured with respect to the C'β atom following the glycine. Thus, glycine residues contribute to the local backbone twist indirectly, by being included in the virtual bond Cα–Cα that spans from the residue preceding the glycine to that which follows it.

Backbone twist profiles (plots of 9 as a function of the amino acid residue) serve to characterize polypeptide chain conformations. Certain conformational characteristics of polypeptides are more clearly seen using 9 values instead of the $\varphi\psi$ values for individual residues. In our plots, the value of the torsion angle Cα–Cβ–C'α–C'β is assigned to the second (C') residue. The angle 9 is related to "the amount of twist per 2 residues", defined as δ by Chou *et al.* (1982); in fact, $9 = \frac{1}{2}\delta$. It thus follows that 9 can be obtained from the helical parameters n (number of residues per turn), h (the rise per residue) and T ($T = 360°/n$) in a corresponding way to that described for δ by Chou *et al.* (1982).

3. Results

(a) Domain–domain contact surfaces

We identified the residues that form the interface between VL and VH by calculation of the solvent-accessible surface of the domains, first in isolation and second when associated. Any residue that lost surface on the association of VL and VH was taken as part of the interface between them. We also determined which residues form van der Waals' contacts across the interface (distance cutoff 4·1 Å). The lists of residues obtained by the two methods were very similar. Thus, except for a few marginal cases, the residues that lose surface in domain-domain contacts also have van der Waals' interactions between the domains, indicating that the VL–VH interface is tightly packed.

The total surface areas of the separated VL and VH domains and that buried on the association is shown in Table 2. The values for the buried surface area (between 1700 and 1900 Å²) and the fraction of the buried surface that is composed of polar atoms are similar to those found in other cases (Chothia & Janin, 1975). For the bovine pancreatic trypsin inhibitor and trypsin it is known that the structure of the isolated proteins does not change significantly on association. In most cases, as for the VL and VH domains considered here, there are no data concerning the structure of the unassociated domains.

Of the total area buried between the VL–VH dimers about one quarter comes from residues in the hypervariable regions and about three quarters from residues in β-sheets. Figure 2 shows the residues that form the interfaces and the areas that are buried for the three VH–VL packings. Two important features are evident in this Figure. First, homologous residues form the interface in the three structures. Second, the pattern formed by the contact residues is most unusual. The contacts of residues on the edge strands of the β-sheets are more extensive than those of residues on the inner strands. This is the opposite of the behavior found in previously described β-sheet packings, where it is the central strands that have the largest contact.

For example, for packing of β-sheets in the same domain, the region of maximal contact generally runs diagonally across the sheets at 45° with respect to the β-strands (Cohen et al., 1981b; Chothia & Janin, 1981). The point is clearly illustrated in the Cα backbone plot in Figure 2(c); here, for each of the Cα atoms a circle is displayed, the area of which is proportional to the total contact area made by the residue with the other sheet. As we describe below, the unusual packing is a direct consequence of the distortions present in this type of β-sheet.

### (b) Conformation of interface β-sheets

The deviation of the conformations of the β-sheets that form the interface between VL and VH from the idealized flat structure (i.e. twisting, coiling and bending) can be characterized by the variations in the twist angle 9 (see Materials and Methods). On such twist profiles, regular twisted β-sheets correspond to horizontal lines with an average $9 = +20°$, right-handed α helices to lines of $9 = -110°$ and tight reverse turns as triplets of points of approximately the same magnitude and alternating sign. The insertion of an additional residue in an edge strand of a β-sheet, so that two edge residues face one another on an inner strand, forms what has been called a β-bulge (Richardson et al., 1978). Such insertions can have a variety of conformational effects depending upon the exact $\varphi\psi$ values of the inserted residue and those of its neighbors. Usually a sharp bend or local coiling is produced in the edge strand; this gives rise to a single- or double-point peak or trough in the 9 values.

In Figure 3 we show the 9 values for the VL–VH interface segments (β-strands with the adjacent hypervariable loops) in KOL, NEW and MCPC 603. Two important features of these β-sheets are evident from the Figure. First, most of the individual values of 9, and the patterns formed by the variations in 9 angles, are very similar in the different sheets, particularly in the inner β-strands ($\beta1, \beta3, \beta5$ and $\beta8$ of Fig. 3) and in the β-bulges; the edge β-strands ($\beta2, \beta4, \beta6$ and $\beta9$ of Fig. 3) have

Table 2

Accessible surfaces and those lost on VL–VH association (Å²)

Domain pair	Isolated surface			Contact surface		
	Hydrophobic	Polar	Total	Hydrophobic	Polar	Total
KOL VL domain	1121	658	1779	580	311	891
KOL VH domain	1216	700	1926	615	250	865
VL–VH in KOL	2337	1358	3705	1195	561	1756
NEW VL domain	1233	744	1977	529	387	916
NEW VH domain	1186	801	1985	506	386	892
VL–VH in NEW	2419	1545	3962	1035	773	1808
MCPC 603 VL domain	1082	689	1771	676	299	975
MCPC 603 VH domain	1156	760	1916	619	324	943
VL–VH in MCPC 603	2238	1449	3687	1295	623	1918
VL–VH average	2331	1714	3785	1195	652	1827

VL

(a)

VH

(b)

(c)
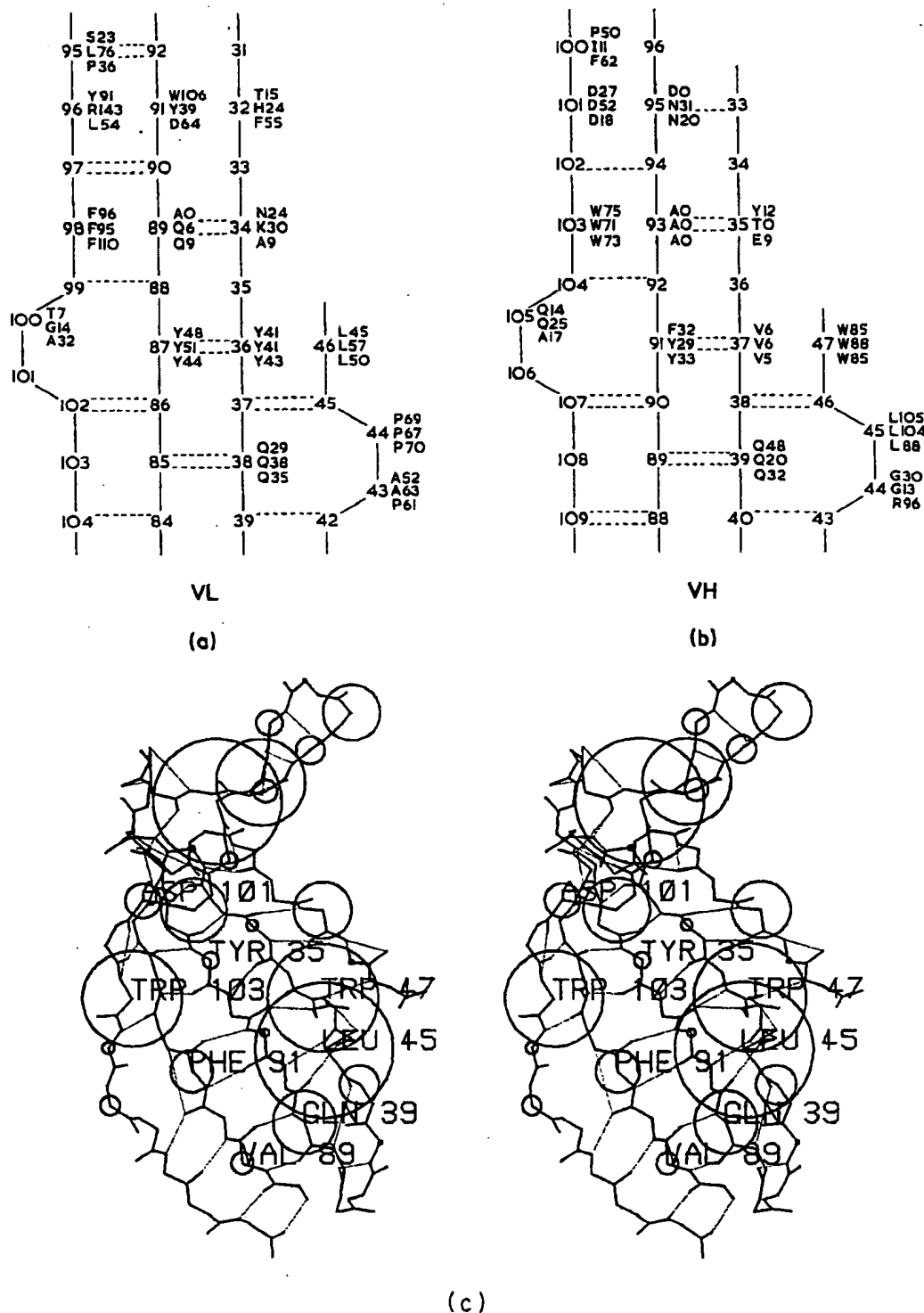
Figure 2. β-Sheet residues that form the VL–VH interface in the Fabs KOL, NEW and MCPC 603. Residue numbers are those of Kabat *et al.* (1983). (a) VL interface-forming β-sheet; (b) VH interface-forming β-sheet. Broken lines indicate hydrogen bonds. At each position where a residue forms part of the interface, we give the residue identity in KOL, NEW and MCPC 603, and the accessible surface of the residue that is buried in the VL–VH interface. Note the β-bulges in the edge strands at positions 43, 44 and 100, 101 in VL and 44, 45 and 105, 106 in the VH. (c) The β-sheet from KOL VH domain. Residues making contacts to the VL domain across the domain–domain interface are circled. The main-chain atoms are displayed. The circles associated with each Cα atom have an area proportional to the accessible surface area lost when the VL–VH dimer forms. Note the large areas associated with residues in the edge strands of the β-sheet.
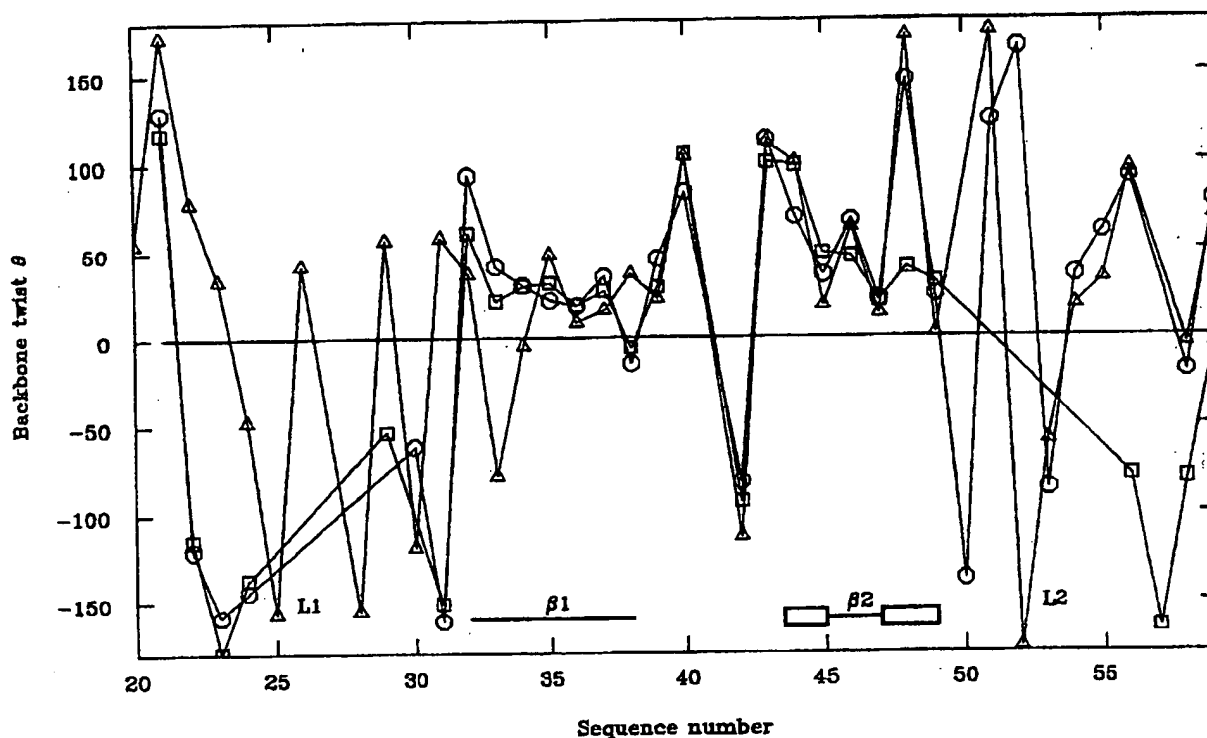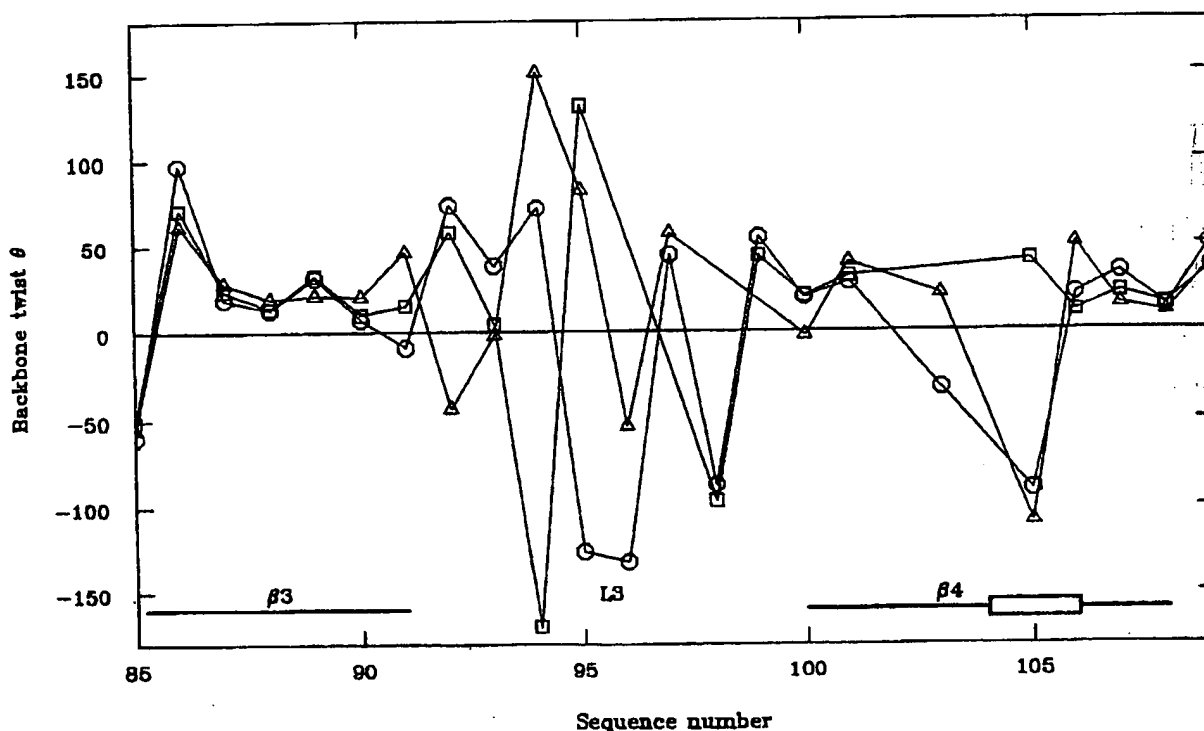
Figure 3. The backbone twist (θ) profiles of VL–VH interface-forming segments. The segments shown include the hypervariable loops (L1, L2, L3, H1, H2 and H3) and the β-strands. The β-strands are indicated by bars at the bottom of the plots and labeled β1 through β9 according to Novotný *et al.* (1983). β-Bulges are denoted by open boxes. Sequence numbers correspond to the Kabat *et al.* (1983) numbering system and are the same as in Fig. 2. (a) and (b) The 2 interface-forming segments of the VL domain: (c) and (d) the 2 interface-forming segments of the VH domain. (O) KOL; (□) NEW; (△) MCPC 603.

Backbone twist θ

H1 $\beta5$ $\beta6$ H2

Sequence number

(c)

Backbone twist θ

$\beta8$ H3 $\beta9$

Sequence number

(d)
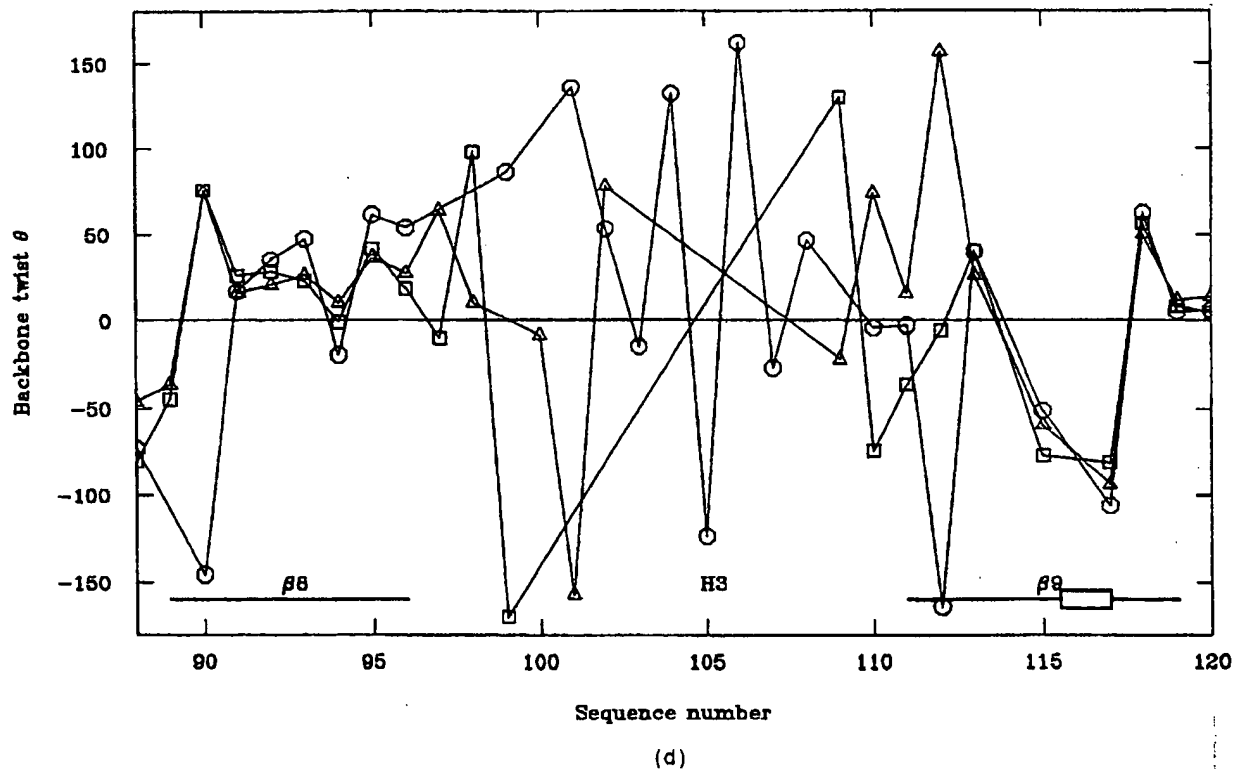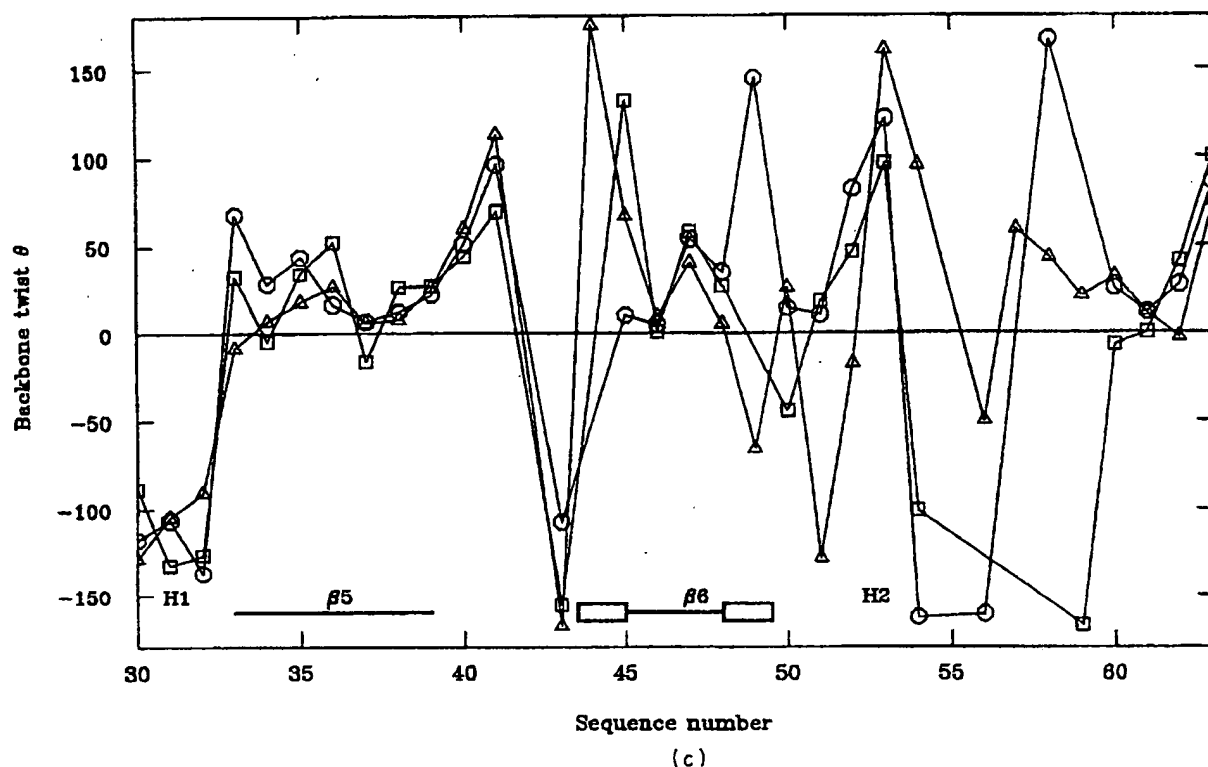
Fig. 3.

greater differences. Conservation of β-bulge conformations is especially striking and implies that they are important architecturally, as previously suggested by Richardson (1981). The correspondence in the β-sheets is made even more evident by the difference in behavior of the hypervariable loops. The overall similarity of β-sheet geometries is confirmed by a least-squares fit of their atomic co-ordinates. Fits of the main-chain atoms of the three VL β-sheets to each other
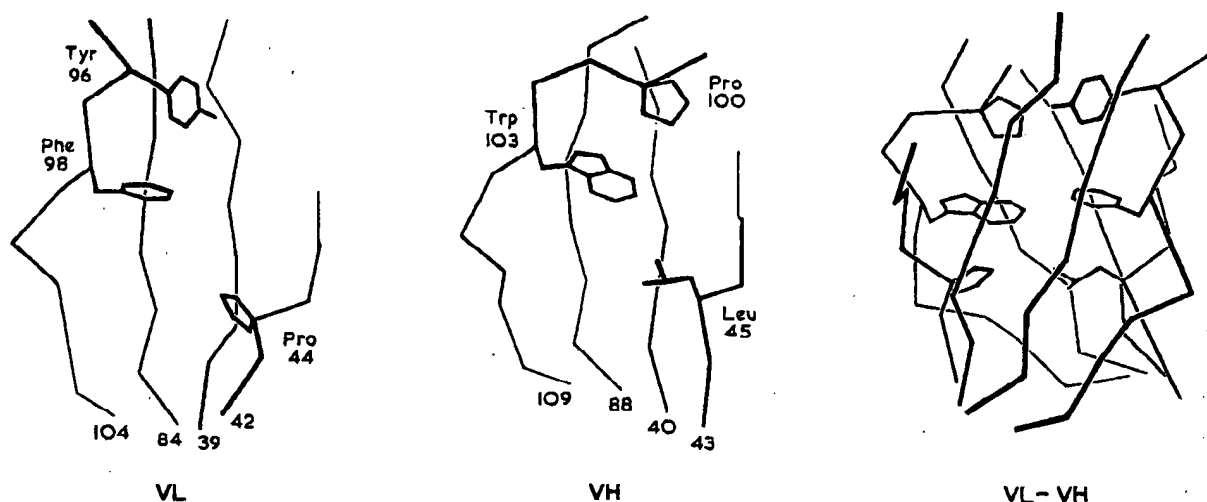
Figure 4. The key residues in the edge strands involved in VL–VH packings (Fab KOL). Note how in (a) Pro44, Tyr96 and Phe98 in VL and in (b) Leu45, Pro100 and Trp103 in VH fold over the central strands of their β-sheets and so in (c) form the core of the VL–VH packing (see also the position of these residues in Fig. 5).

(30 residues), and of the three VH β-sheets to each other (32 residues) give root-mean-square differences in atomic positions of between 0·73 and 1·23 Å (see Table 3). If a few peripheral residues are removed from the fits the r.m.s.† differences are reduced to 0·55 to 0·87 Å. Table 3 also reports the results of least-squares fits of the VL β-sheets to the VH β-sheets. The r.m.s. differences are only a little greater than for the fits of the VL or VH β-sheets to each other, 0·70 to 1·11 Å. Thus, the six regions of β-sheet that form the VL–VH interface in KOL, NEW and MCPC 603 have very similar structures. In fact, the least-squares superposition of the two sheets can be achieved as a 2-fold symmetry operation, i.e. rotation around an axis passing through the centroid of the interface.

The second feature of the β-sheets illustrated in Figure 3 is the different amounts of twist found in the edge and inner strands. The two central strands in both VL and VH have ϑ values in the range that indicate a degree of twist commonly found in β-sheets. The average ϑ value tends to be the same for both the inner and the edge strands, but the twist of the edge strands is dominated by β-bulges (Figs 2 and 3) with characteristic ϑ values ±70. Its effect is to fold the ends of the edge strands over central strands. This occurs at two diagonally opposite corners of the β-sheets. Side-chains of residues 44 (Pro), 96 (Tyr, Arg, Leu) and 98 (Pro) in VL and 45 (Leu), 100 (Pro, Ile, Phe) and 103 (Trp) cover residues in the inner strands (Fig. 4(a) and (b)). The other parts of the edge strand residues, 45–46 and 101–104 in VL, 46–48 and 106–109 in VH, lie next to the inner strand in the normal manner.

(c) Packing of the β-sheets at the VL–VH interface

As noted above, the strong twists that occur in the edge strands of VL and VH means that residues at two diagonally opposite corners fold over the β-sheets: 44, 96 and 98 in VL (Fig. 4(a)), and 45, 100 and 103 in VH (Fig. 4(b)). Figure 4(c) shows that when the VL and VH domains pack together

Table 3

The fit of β-sheets forming VL–VH interfaces

A. *Fits of individual β-sheets*

	VL†			VH‡		
	KOL	NEW	MCPC	KOL	NEW	MCPC
VL† KOL	—	0·76	0·55	0·88	1·11	0·94
NEW	—	—	0·82	0·96	1·05	0·97
MCPC			—	0·70	1·00	0·84
VH‡ KOL				—	0·87	0·65
NEW					—	0·87
MCPC						—

B. *Fits of both β-sheet regions of the VL–VH interfaces§*

	KOL	NEW	MCPC
KOL	—	0·87	0·70
NEW		—	0·87
MCPC			—

The Table gives r.m.s. differences in position of the main chain atoms following least-squares fits of their co-ordinates. Differences are given in Å.

† VL residues used to determine fits and r.m.s. differences 33–39, 43–47, 84–90 and 98–104.

‡ VH residues used to determine fits and r.m.s. differences 33–40, 44–48, 88–94 and 102–109.

§ Residues used in fits 33–39, 43–47, 84–90 and 98–104 of VL and 34–40, 44–48, 88–94 and 103–109 of VH.

† Abbreviation used: r.m.s., root-mean-square.

(a)

(b)

(c)

(d)
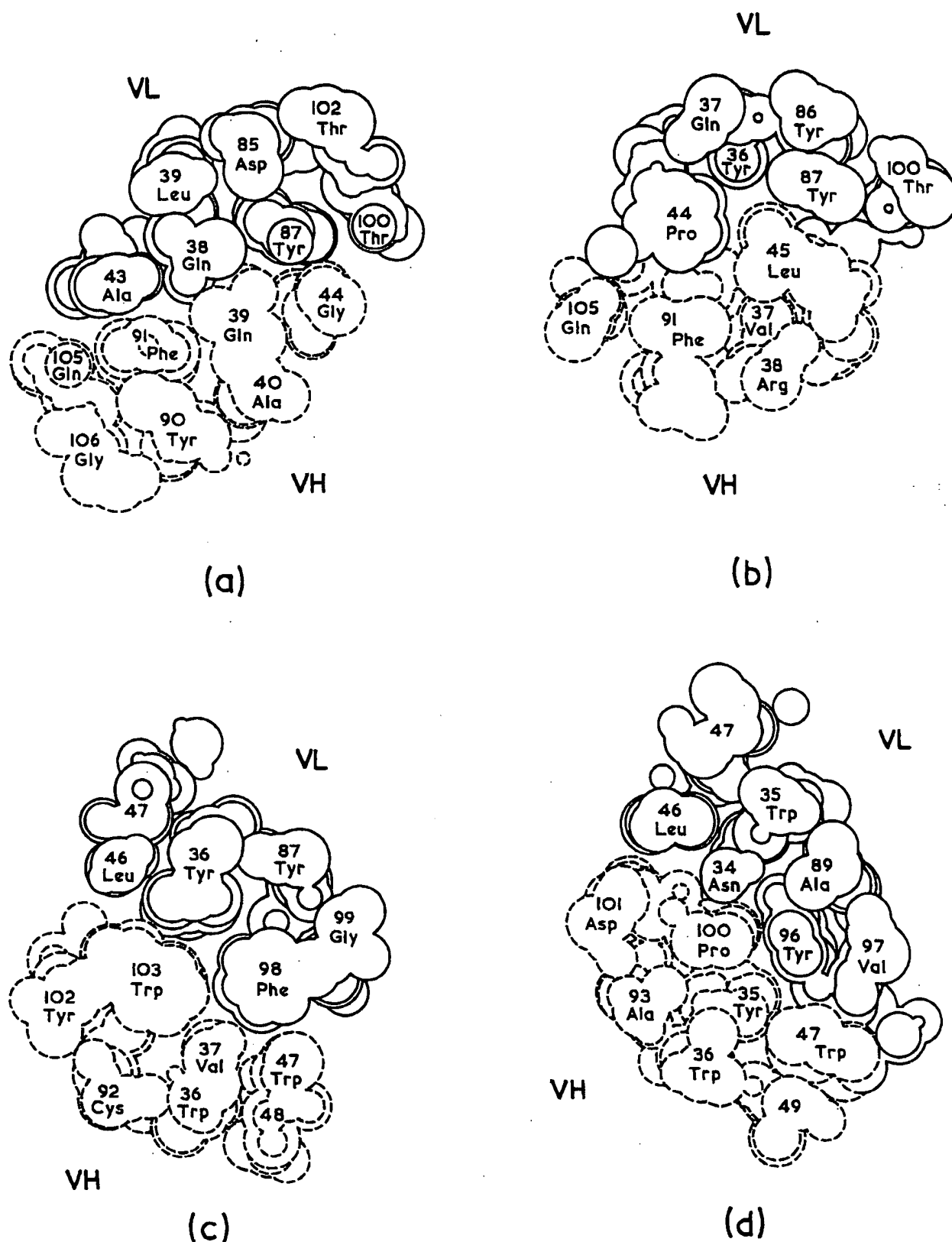
Figure 5. Residue packing at the KOL VL–VH interface.. This Figure shows superimposed serial sections cut through a space-filling model of the interface. VH residues are shown by broken lines and VL residues by continuous lines. The pseudo 2-fold axis that relates VL to VH is perpendicular to the page. Each part of the Figure shows 4 sections, separated by 1 Å, superimposed. (a) Sections 0 to 3 Å; (b) sections 4 to 7 Å; (c) sections 8 to 11 Å; and (d) Sections 12 to 15 Å.

Table 4

Residues buried in VL–VH interfaces

Domain	Residue No.	Residue at this position in			Accessible surface area of residue (Å²)			No. of sequences known that include this position†	Principal residues found at this position† (identity and number of cases)
		KOL	NEW	MCPC	KOL	NEW	MCPC		
VL	34	Asn	Lys	Ala	2	39	0	362	Ala117, Asn92, His51, Ser37
	36	Tyr	Tyr	Tyr	0	0	1	318	Tyr243, Phe40, Val28
	38	Gln	Gln	Gln	2	7	17	302	Gln279
	44	Pro	Pro	Pro	8	5	5	238	Pro190, Phe29, Val14
	46	Leu	Leu	Leu	17	35	8	235	Leu157, Gly32, Pro19, Val13
	87	Tyr	Tyr	Tyr	9	1	11	227	Tyr160, Phe65
	89	Ala	Gln	Gln	0	1	0	217	Gln128, Ala35
	91	Trp	Tyr	Asp	3	12	0	211	Trp59, Tyr31, Ser27
	96	·Tyr	Arg	Leu	6	5	3	199	Trp46, Tyr31, I26, R20
	98	Phe	Phe	Phe	10	9	2	206	Phe203
VH	35	Tyr	Thr	Gln	0	2	0	217	Gln53, Asn42, Ser34, Lys22
	37	Val	Val	Val	0	3	1	200	Val178, Ile19
	39	Gln	Gln	Gln	8	20	21	183	Gln176
	45	Leu	Leu	Leu	10	6	3	163	Leu160
	47	Trp	Trp	Trp	11	6	4	157	Trp151
	91	Phe	Tyr	Tyr	0	8	11	159	Tyr128, Phe30
	93	Ala	Ala	Ala	0	0	1	161	Ala146
	95	Asp	Asn	Asn	0	0	5	131	Asp53, Gly18
	100	Pro	Ile	Phe	0	32	0	113	Phe76, Met11, Leu6
	103	Trp	Trp	Trp	27	28	26	125	Trp118

† Data taken from Kabat et al. (1983).

these six residues form the center of the interface. They are in contact with each other in pairs and make a herringbone pattern.

Details of how residues pack at the VL–VH interface can be seen in sections cut through space-filling models. Figure 5 shows sections of the KOL VL–VH interface. The central role played by the three pairs of edge residues, Tyr96 and Trp103, and Leu45 and Pro44 are seen in parts (b), (c) and (d) of the Figure. The inner strands of the β-sheets, 32–39 and 84–92 in VL and 33–40 and 88–95 in VH, only make interdomain contacts at one end of the interface where the side-chains of Gln38 and Gln39 hydrogen bond to each other (Fig. 5(a)). The structures of the VL–VH interfaces in NEW and MCPC 603 are very similar to that of KOL illustrated here. This is demonstrated by graphical inspection of their packing and by the fits of the co-ordinates of the main-chain atoms forming the VL–VH interfaces described above (Table 3).

The packing of the β-sheets at the three VL–VH interfaces can be described in terms of a three-layer structure: an inner layer consisting of large side-chains from strongly twisted ends of the edge strands; and two outer layers formed by the main and side-chains of the inner β-strands and the middle part of the edge strands (Figs 4 and 5).

(d) Three-layer packing as a general model for VL–VH associations

Ten years ago Poljak et al. (1975) examined their Fab NEW structure and noted that the residues

that form the VL–VH interface were conserved in the other immunoglobulin sequences then known. They predicted that the mode of association of other VL–VH dimers would be the same as that found in Fab NEW. The structures and many sequences determined since then, and the work reported here, confirm their prediction.

The three structures studied here include a wide range of immunoglobulins: human λ and γ to mouse κ and γ (Table 1). In KOL, NEW and MCPC 603 residues at about ten positions in VL and in VH are buried in the interface between the domains. The amino acid sequences of many other immunoglobulins have been determined and a tabulation published by Kabat et al. (1983). We examined the tables of VL and VH sequences to find what residues occur at positions homologous to the 20 buried in the VL–VH interfaces studied here. The results of this survey are given in Table 4 and Figure 6.

At 12 of the 20 positions residue identity is absolutely, or very strongly, conserved: in VL, residues 36, 38, 44, 87 and 98; in VH, residues 37, 39, 45, 47, 91, 93 and 103. As shown in Figure 6, these residues form the whole of the central and lower regions of the interface. The eight positions that have some variation in residue identity are all in the upper part of the interfaces where they are adjacent to and partially buried by the hyper-variable regions. The three structures studied have a range of residues at these positions that is fairly representative of those found in other sequences (Table 4). Inspection of the three structures shows

VL

96 0·23 W	91 0·28 W	
0·16 Y	0·25 Y	
98 0·99 F	89 0·60 Q	34 0·32 A
	0·16 A	0·25 N
99 1·00 G		
	87 0·99 F/Y	36 0·90 Y/F 46 0·67 L
		0·14 G
101 1·00 G		
		38 0·92 Q 44 0·80 P
		0·12 F

VH

100 0·73 F		
0·10 L	95 0·40 D	
	0·14 G	
103 0·94 W	93 0·91 A	35 0·23 E
		0·19 N
104 0·98 G		
	91 0·99 Y/F	37 0·89 V 47 0·93 W
		0·10 I
106 0·98 G		
		39 0·96 Q 45 0·98 L

Figure 6. The conservation of residues that form VL–VH interfaces. On a plan of the VL and VH β-sheets we show the principal residues found at sites buried in the interface and at sites involved in the formation of the β-bulges. At each site we note the proportion of known sequences that contain the given residue, for example, 0·99 of known VL sequences have Phe at position 98 (see Table 4). The one-letter code for amino acids is used.

that the different residues are accommodated by small conformational changes in the ends of the β-strands and somewhat larger changes in the hypervariable regions. The Gly–X–Gly sequence that produces the β-bulge at residues 99–101 in VL and 104–106 in VH is absolutely conserved in VL and VH sequences (Fig. 6). Thus the pattern of conserved residues in VL and VH sequences suggests that the three-layer packing found for the structures studied here is a general model for VL–VH associations.

(e) Three-layer packing as a new β-sheet packing class

The packing of β-sheets in proteins has been analyzed in some detail (Chothia *et al.*, 1977; Chothia & Janin, 1981, 1982; Cohen *et al.*, 1981b). Observed packings have been divided into two classes: the aligned and the orthogonal. The three-layer packings described here for the VL–VH
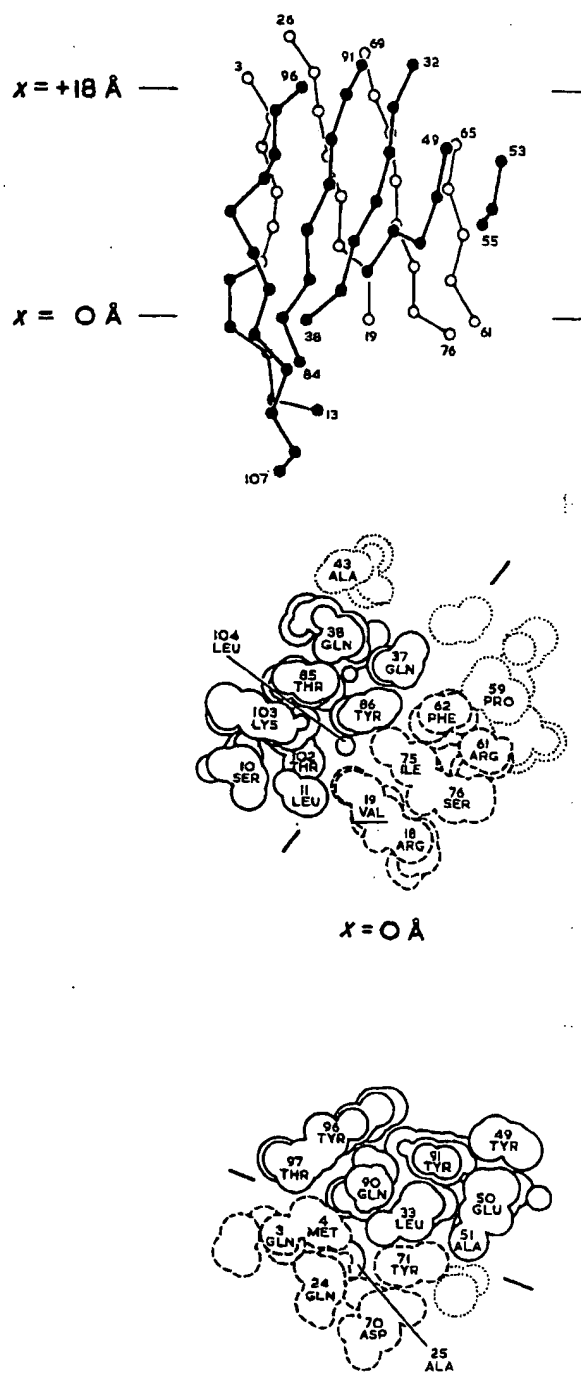
Figure 7. An example of the aligned class of β-sheet packings. This Figure shows the packing of 2 β-sheets within a VL domain. (a) Shows arrangement of the β-sheets: Cα atoms in one sheet are indicated by open circles and those in the other sheet by filled circles. Sections cut through a space-filling model of the packing at x = 0 Å and x = 18 Å are shown in (b) and (c). In (b) and (c) the strands of the sheet are approximately perpendicular to the page. Note how the strands of one β-sheet make direct contact with the strands in the other sheet. Residues from the edge strands do not form a middle layer as they do in the 3-layer packings illustrated in Figs 4 and 5. Adapted from Chothia & Janin (1981).

interface do not fit into either of these classes. Orthogonal β-sheet packings are quite different: in that class the main chain directions of the packed sheets are inclined at approximately 90° and one or more strands pass from one β-sheet to the next with little or no interruption (Chothia & Janin, 1982).

The aligned packings do have some similarities to three-layer packings in that both involve β-sheets that are essentially independent. The packing of residues at the interface, however, is very different. In aligned packings the β-sheets pack face-to-face with some of the side-chains of each strand making direct contact with strands on the opposite sheet. Examples of such packings are found within each of the VL and VH domains (Fig. 7). Aligned β-sheet packings can be described as two-layer structures with the side-chains of the two layers packed together. Insertion between the two sheets of the side-chain from a residue of an edge strand is uncommon and where it does occur is only found at the margins of the interface. In the three-layer packings described here the residues from the edge strands form a complete layer at the center of the interface (Fig. 5).

The different residue packing in the two classes results in different geometry. In aligned packings the angle between the mean chain directions of the packed β-sheets is about $-30°$ ($-20°$ to $-50°$) and arises from the twist of the individual β-sheets. The angle of the three-layer packings described here is $-50°$. This angle arises from the bulkier size of the residues that form the middle layer, as well as from the concave curvature of the interface-forming β-sheets (cf. Fig. 1). In aligned packings the β-sheets are about 10 Å apart. In three-layer packings the central regions of each sheet are 14 Å apart; the larger value arises from the additional central layer. This central layer extends through the whole length of the interface, from the "bottom" up to the "top" where the binding site is located, and participates in forming the floor of the antigen combining cavity. For example, some of the aromatic third-layer side-chains (Phe98 in the VL domains) were shown to be indispensable for the antigenic specificity (Azuma et al., 1984), even though they are only marginally exposed to solvent (Novotný et al., 1983).

The general shape of the aligned β-sheet packing is that of a twisted prism (Fig. 7; Chothia & Janin, (1981). If we ignore the regular parts of the edge strands, three-layer packings have the general shape of a twisted hyperboloid with elliptical cross-sections (Figs 4 and 5; Novotný et al., 1983, 1984).

Important for the three-layer packings described here are the aromatic side-chains that form the center of the contact. They pack with their side-chains approximately perpendicular to each other as typified in benzene crystals (Cox et al., 1958; Wyckoff, 1969; Nockolds et al., 1975; Thomas et al., 1982; Williams, 1980; Burley & Petsko, 1985; Novotný & Haber, 1985). This contrasts with the residues most commonly found at the interface of aligned packings. They are aliphatic residues like

Val, Ile and Leu that approximate the close packing expected for hard spheres.

4. Conclusion

Immunoglobulins are composed of sets of domain dimers. Single domains are sandwiches composed of two β-sheet backbone layers with the hydrophobic side-chains in between. For the VL–VH domain dimers we have found that they cannot be described simply as the packing together of two of these sandwich structures. Instead, the VL–VH interface has a three-layer structure with a set of primarily aromatic side-chains interposed between the sandwich structure making up each of the domains. The three-layer packing is facilitated by highly twisted edge strands that bend at places where β-bulges occur. This mode of β-sheet packing is different from those described previously and produces a sheet–sheet interface that is significantly bulkier than typical "aligned" sheet–sheet interfaces (Chothia & Janin, 1981). The interfaces between VL and VH in Fab KOL, Fab NEW and Fab MCPC 603 have very similar structures of this three-layer type. The pattern of residue conservation found in the sequences of other immunoglobulins strongly suggests the same structure occurs generally in VL–VH association. This is in accord with the presence of β-bulges at homologous postions in the edge β-strands, and their highly conserved conformation. The three-layer β-sheet packing thus plays a central role in forming the antibody combining site.

References

Azuma. T., Igras, V., Reilly, E. B. & Eisen. H. (1984). Proc. Nat. Acad. Sci., U.S.A. 81, 6139–6143.

Beale, D. & Feinstein, A. (1976). Quart. Rev. Biophys. 9, 135–180.

Bernstein, F. C., Koetzle. T. F., Williams, G. J. B., Meyer. E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). J. Mol. Biol. 112, 535–542.

Brooks. B.. Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). J. Comput. Chem. 4, 187–217.

Burley, S. & Petsko, G. (1985). Science, 229, 23–28.

Chothia. C. (1973). J. Mol. Biol. 75, 295–302.

Chothia, C. & Janin, J. (1975). Nature (London), 256, 705–708.

Chothia, C. & Janin, J. (1981). Proc. Nat. Acad. Sci., U.S.A. 78, 4146–4150.

Chothia, C. & Janin, J. (1982). Biochemistry, 21, 3955–3965.

Chothia. C.. Levitt, M. & Richardson, D. (1977). Proc. Nat. Acad. Sci., U.S.A. 74, 4130–4134.

Chou. K. C., Pottle, M., Neméthy, G., Ueda, Y. & Scheraga, H. (1982). J. Mol. Biol. 162, 89–112.

Cohen, F. C., Sternberg, M. J. E. & Taylor, W. R. (1980). Nature (London), 285, 378–382.

Cohen, F. C., Novotný, J., Sternberg, M. J. E., Campbell. D. G. & Williams. A. F. (1981*a*). *Biochem. J.* **195**. 31–40.

Cohen, F. C., Sternberg, M. J. E. & Taylor, W. R. (1981*b*). *J. Mol. Biol.* **148**, 253–272.

Cox, E. G., Cruickshank, D. W. J. & Smith, J. A. S. (1958). *Proc. Roy. Soc. ser. A* **247**, 1–21.

Cunningham, B. A., Wang, J. J., Berggard, I. & Peterson. P. A. (1973). *Biochemistry*, **12**, 4811–4816.

Deisenhofer, J. (1981). *Biochemistry*, **20**, 2361–2370.

Edmundson, A. B., Ely, K. R., Abola, E. E., Schiffer. M. & Panagiotopoulos, N. (1975). *Biochemistry*. **14**. 3953–3961.

Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M. & Huber, R. (1974). *Eur. J. Biochem.* **45**, 513–524.

Feinstein, A. (1979). *Nature (London)*, **181**, 230.

Goto, Y. & Hamaguchi, K. (1982). *J. Mol. Biol.* **156**, 911–926.

Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I. & Davis, M. M. (1984). *Nature (London)*, **308**, 145–149.

Hochman, J., Inbar, D. & Givol, D. (1973). *Biochemistry*, **12**, 1130–1135.

Jensenius, J. C. & Williams, A. F. (1982). *Nature (London)*. **300**, 583–588.

Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Milner, M. & Perry, H. (1983) *Sequences of Proteins of Immunological Interest*, Public Health Service. N.I.H., Washington, D.C.

Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

Marquart, M., Deisenhofer, J. & Huber, R. (1980). *J. Mol. Biol.* **141**, 369–391.

Mostov, K. E., Friedlander, M. & Blobel, G. (1984). *Nature (London)*, **308**, 37–43.

Nockolds, C. E., Kretsinger, R. H., Coffee, C. J. & Bradshaw, R. A. (1975). *Proc. Nat. Acad. Sci.*, *U.S.A.* **69**, 581–584.

Novotný, J. & Auffray, C. (1984). *Nucl. Acid Res.* **12**, 243–255.

Novotný, J. & Franěk, F. (1975). *Nature (London)*, **258**, 641–643.

Novotný, J. & Haber, E. (1985). *Proc. Nat. Acad. Sci.*, *U.S.A.* **82**, 4592–4596.

Novotný, J., Vitek, A. & Franěk, F. (1977). *J. Mol. Biol.* **113**. 711–718.

Novotný, J., Bruccoleri, R., Newell, J., Murphy, D., Haber, E. & Karplus, M. (1983). *J. Biol. Chem.* **258**. 14433–14437.

Novotný, J., Bruccoleri, R. & Newell, J. (1984). *J. Mol. Biol.* **177**. 567–573.

Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro. J. & Strominger, J. L. (1979). *Nature (London)*. **282**. 266–270.

Pauling, L., Corey, R. B. & Branson, H. R. (1951). *Proc. Nat. Acad. Sci.*, *U.S.A.* **37**, 205–211.

Phizackerley, R. P., Wishner, B. C., Bryant, S. H., Amzel, L. M., Lopez de Castro, J. A. & Poljak, R. J. (1979). *Mol. Immunol.* **16**, 841–850.

Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. (1975). *Immunogenetics*, **2**. 393–394.

Richmond, T. J. & Richards, F. M. (1978). *J. Mol. Biol.* **119**, 537–555.

Richardson, J. S., Getzoff, E. D. & Richardson, D. C. (1978). *Proc. Nat. Acad. Sci.*, *U.S.A.* **75**. 2574–2578.

Richardson, J. S. (1981). *Advan. Protein Chem.* **34**, 167–339.

Saul, F. A., Amzel, L. M. & Poljak, R. J. (1978). *J. Biol. Chem.* **253**, 585–597.

Segal, D., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. (1974). *Proc. Nat. Acad. Sci.*, *U.S.A.* **71**, 4298–4302.

Schiffer, M., Girling, R. L., Ely, K. R. & Edmundson, A. B. (1973). *Biochemistry*, **12**, 1620–1631.

Shrake, H. & Rupley, J. A. (1973). *J. Mol. Biol.* **79**, 351–371.

Thomas, K. A. Smith, G. M., Thomas, T. B. & Feldmann. R. J. (1982). *Proc. Nat. Acad. Sci.*, *U.S.A.* **79**, 4843–4847.

Williams, D. E. (1980). *Acta Crystallogr. sect. A*, **36**, 715–723.

Wyckoff, R. W. G. (1969). In *Crystal Structures*. 2nd edit., vol. 6, part 1. pp. 1–2, Wiley, New York.

Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S. P., Aleksander, I. & Mak, T. W. (1984). *Nature (London)*, **308**, 145–149.

Edited by R. Huber

Structural invariants of antigen binding: Comparison of immunoglobulin V_L–V_H and V_L–V_L domain dimers

(antibody/β-sheet/electrostatics/strophoid/x-ray structure)

JiŘí Novotný and Edgar Haber

Molecular and Cellular Research Laboratory, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114

Communicated by Elkan R. Blout, March 11, 1985

ABSTRACT   Antigen-combining site arises by noncovalent association of the variable domain of the immunoglobulin heavy chain (V_H) with that of the light chain (V_L). To analyze the invariant features of the binding region (V_L–V_H domain interface), we compared the known immunoglobulin three-dimensional structures by a variety of methods. The interface forms a close-packed, twisted, prism-shaped "β-barrel" characterized by cross-sectional dimensions 1.04 × 0.66 nm and a top-to-bottom twist angle of 212°. The geometry of the interface is preserved via invariance of some 15 side chains, both inside the domains and on their surface. Buried polar residues form a conserved hydrogen-bonding network that has a similar topological connectivity in the two domain types; two hydrogen bonds contributed by invariant side chains extend across the interface and anchor the β-sheets in their relative orientation. Invariant aromatic residues close-pack at the bottom of the binding-site β-barrel with their ring planes oriented perpendicularly in the characteristic "herringbone" packing mode. Electrostatic computations that implicitly include solvent effects show the domains to be stabilized by large electrostatic forces. However, structures that were crystallized at lower pH have their electrostatic energies appropriately lowered, implying that full ionization of carboxyl side chains is essential for efficient electrostatic stabilization. The unusual mode of domain–domain association in the V_L–V_L dimer RHE correlates with its overall repulsive electrostatic energy (+54 kJ/mol), as opposed to negative (i.e., stabilizing) energy values (−263 to −543 kJ/mol) found in the domains of the other structures. The V_L–V_L dimer REI mimics closely the interface geometry of V_L–V_H dimers although its domain–domain contact area is lower by 18%.

The structural diversity of antibody molecules epitomizes one of the most interesting problems of molecular biology, that of a relationship between molecular shape and biological function. Although spatial structures of three different antibody binding sites have been elucidated (1–3), our knowledge of structural prerequisites of the antigen binding function remains incomplete. Relative importance of individual amino acid residues that form the antigen binding site is not well understood, nor is it clear how their presence influences antigen binding. Recent advances in genetic engineering have put at our disposal a means to modulate antibody specificity at will—e.g., by applying site-directed mutagenesis to genes encoding immunoglobulin polypeptide chains. This possibility, however, can only be realized on the basis of a sound understanding of principles that determine protein anatomy in general (4, 5) and that of antibody molecules in particular (6).

The antigen-combining site is formed by noncovalent association of two "variable" (V) domains provided by two different polypeptide chains, heavy (H) and light (L). The V_L–V_H interface consists of two closely packed β-sheets and its geometry corresponds to a nine-stranded elliptical (4) or prism-shaped (7) barrel. The barrel forms the bottom and sides of the antigen binding site, and amino acid residues that are part of the domain–domain interface and appear not to be accessible to solvent or antigen contribute to antibody specificity (6).

Here we study those conformational features of the V_L and V_H domains that are conserved in all the antibodies and form the constant scaffold for the binding site. We do so by comparing three-dimensional structures with use of a novel procedure (8, 9): we superimpose, by the least-squares method, only those side chain atoms that are invariant in all the immunoglobulins. This allows for differences in functional importance of different parts of the structure and leads to more meaningful results than the method employed previously (10, 11), namely, a least-squares superposition of complete polypeptide chain backbones (i.e., optimization of structural correspondence over the domain as a whole). We also analyze the conserved hydrogen-bonding network existing among polar side chains that are buried inside the domains and discuss the contribution of electrostatic interactions to the stability of the binding site.

ATOMIC COORDINATES AND CALCULATIONS

Crystallographic coordinates of human Fab fragments NEW, KOL, and MCPC 603; V_L–V_L dimers RHE and REI; and Bence Jones protein (light-chain dimer) MCG were obtained from the Brookhaven Protein Data Bank (12). No attempts were made to energy-minimize or otherwise improve the original data. Structural manipulations, such as least-squares superpositions, generation of hydrogen bonded lists, potential energy evaluations, etc., were performed with the program CHARMM version 16 (13) as described (6, 14). The electrostatic potential was computed by use of a solvent-modified Coulomb formula (14). The effect of solvent was modeled by multiplying charges on atoms by a constant that depends linearly on the distance of the atom from the surface of the protein (15). The potential was evaluated to infinity; i.e., no distance cutoff was applied to evaluations of pairwise atomic interactions. Stereo drawings were made from CHARMM-generated files, using previously described graphics facilities and software (14). Amino acid alignments of immunoglobulin variable domains used as a basis for structural comparisons and residue numbering were those of Kabat *et al.* (16).

RESULTS AND DISCUSSION

Conservation of the Binding Site Geometry. Although all the immunoglobulin domains share the same folding scheme—two antiparallel β-sheets packed face-to-face (4)—the number of β-strands, strand orientation, side chain preponder-

Abbreviations: V_L, light chain variable; V_H, heavy chain variable.

ance, and other structural characteristics differ widely among various domain types (V_L, V_H, and constant domains C_L, C_{H1}, C_{H2} and C_{H3}). Only three residues are common to all the domains: two cysteines that form a disulfide bridge between the β-sheets and a tryptophan that packs against them (17). Structural diversity of this kind correlates with the fact that different domain types perform different biological functions, such as antigen binding in V_H and V_L or complement binding in C_{H2}. Identical domain types, on the other hand, might be expected to have many more structural features in common.

To compare x-ray structures of V_L and V_H domains, the best-resolved crystallographic data-set, the Fab fragment KOL, was first chosen as a reference structure and oriented most conveniently in the reference frame of Cartesian coordinates ("barrel orientation" of figure 4 in ref. 6). Second, selected side chains of the V_L or V_H domains of Fab fragments NEW and MCPC 603 were superimposed, independently in each domain, on the corresponding side chains of the reference structure. Because of their invariance and central positions in the domain cores (17), residues Cys-23 and -88, Trp-35, and Phe-62 were chosen in the V_L domain; Cys-22 and -92, Trp-47, and Leu-78 were chosen in the V_H domain.

Fig. 1 displays the superimposed side chains and resulting polypeptide backbone fits of all the three V_L–V_H dimers. No attempt was made to reproduce the exact mode of domain–domain association with our matching procedure, yet polypeptide chain segments involved in the V_L–V_H interface can be seen to overlay virtually exactly (the average root-mean-square difference of these segments among the three structures compared is 0.11 nm). At the same time, significant differences are apparent in backbone conformations of solvent-facing sides of the dimer (root-mean-square differences of 0.5 nm and more). The close match of segments forming the V_L–V_H interface was striking and suggested that despite amino acid variability of both the variable domains (68% of positions vary in the three V_L domains, and 74% in the V_H domains), side chains are conserved in various regions of V_L and V_H primary structures in such a way as to preserve the geometry of the V_L–V_H interface (the antigen-combining region).

Importance of Exposed Nonpolar and Buried Polar Residues. Naturally, the question arises how conservation of side chains in separate domains gives rise to the invariance of the domain–domain interface.

FIG. 2. Conserved features of the antigen-combining region (V_L–V_H domain interface). (A) Comparison of selected invariant side chains in superimposed KOL, MCPC 603, and NEW V_L and V_H domains. The figure shows C^α plot of a polypeptide backbone of the reference structure (V_L–V_H dimer KOL, V_L domain α-carbons represented by circles) together with selected side chains of all three V_L–V_H dimers. In addition to the residues used to produce the least-squares fit (see Fig. 1 and legend), the side chains that superimpose virtually exactly are Gln-6, Val-19, Gln-37, Leu-47, Ile-48, Leu-73, Glu-81, Asp-82, Tyr-86, and Thr-102 in the V_L domains and Leu-4, Gln-6, Leu-20, Phe-29, Arg-38, Glu-46, Asp-86, and Tyr-90 in the V_H domains. (B) A close-up of the antigen-combining region (V_L–V_H interface) showing positions of invariant residues that mediate domain–domain interaction. The interface-forming polypeptide chain segments of KOL are drawn in light lines (V_L domain with circles). Residues shown from the three structures (heavy lines) were not mutually superimposed; rather, the fit was produced as described in the legend to Fig. 1. Note the six aromatic rings in the interface (Tyr-36, Tyr-87, and Phe-98 of V_L and Trp-47, Tyr-91, and Trp-103 of V_H). The two forked side chains at the bottom of the binding site β-barrel are glutamine residues 38 (V_L) and 39 (V_H) involved in interdomain hydrogen bonds. (C) A detailed view of KOL backbone segments that form the interface β-barrel (binding site). Heavy line, β-strands forming the barrel; light line, interstrand hydrogen-bonds and the least-squares-fitted strophoid surface that was used to obtain the dimensions of the barrel (see Table 1 for β-barrel dimensions).

V_H

V_L

FIG. 1. Superposition of KOL, MCPC 603, and NEW V_L and V_H domains. Using the program CHARMM, 4 side chains in each domain were least-squares superimposed on the corresponding side chains from the domain of the reference structure (KOL). The side chains, shown in heavy lines, are the invariant residues Cys-23 and -88, Trp-35, and Phe-62 in the V_L domains and Cys-22 and -92, Trp-36, and Leu-78 in the V_H domains. Polypeptide backbones (light lines) are traced by C^α atoms. Note that polypeptide chain segments involved in the V_L–V_H interface (antigen-combining region) overlay virtually exactly (root-mean-square shift 0.11 nm) despite the fact that the V_L and V_H domains were matched independently and no attempt was made to reproduce the exact mode of domain–domain association.

A

B

FIG. 3.   The conserved hydrogen-bonding pattern provided by polar residues buried inside the V_L and V_H domains of KOL. (A) To facilitate orientation, prominent side chains are displayed and identified by names and numbers in the same orientation as in B. (B) Polypeptide chain backbones of both domains are denoted by heavy lines, and hydrogen bonds by light lines. In addition to the regular interbackbone hydrogen-bonding network characteristic of antiparallel β-sheets, there are hydrogen bonds provided by side-chain atoms. Note the two hydrogen bonds of Gln-38 (V_L) and Gln-39 (V_H) that span the domain–domain interface.

Amino acid alignments show that there are 37 residues conserved among the KOL, NEW, and MCPC 603 V_L domains and 31 residues among the V_H domains. They include the side chains used to produce the least-squares fit, and they always occur at positions invariant in many other V_L and V_H domains (16). The majority of them have hydrophobic side chains, and our solvent-accessibility calculations (18, 19) confirm the previous observation (20) that virtually all of them are buried inside the domains. However, we found that

some of the nonpolar residues are exposed to solvent, while some of the polar ones are buried. The structural importance of the exposed nonpolar and buried polar residues is apparent from the fact that they belong to the most stringently conserved side chains in both V_L and V_H domains (16). The solvent-exposed residues are Tyr-36, Leu-46, Tyr-87, and Phe-98 in the V_L domain and Val-2, Leu-45, Trp-47, and Trp-103 in the V_H domain; the buried residues are Gln-6, Gln-37, Asp-82, and Thr-102 in the V_L domain and Gln-6,

Table 1.   Geometry of domain–domain interfaces (antigen-combining sites)

Structure	Major semiaxis, nm	Minor semiaxis, nm	Helical pitch, nm	Twist angle (top to bottom)	Goodness of fit[*]
KOL V_L–V_H	1.013	0.652	4.195	214°	0.140
MCPC 603 V_L–V_H	1.081	0.662	4.245	220°	0.132
NEW V_L–V_H	1.073	0.630	4.333	210°	0.150
REI V_L–V_L	0.994	0.688	4.270	204°	0.124
Average	1.04 ± 0.04	0.66 ± 0.02	4.26 ± 0.05	212° ± 6°	
MCG V_L–V_L[†]	0.904	0.747	6.223	139°	0.144

It was shown (28) that the geometry of β-sheet–β-sheet interfaces can be approximated by strophoid (twisted hyperboloid) surfaces. The strophoid model gives a significantly better goodness of fit than the cylindrical model of V_L–V_H β-sheet interface (6). The values [nm] were obtained by least-squares fitting of strophoids into polypeptide chain backbones of interface-forming V_L and V_H β-sheets. The surface is mathematically defined by the following four parameters: major and minor semiaxis of (elliptical) cross-section, surface curvature (analogous to the curvature of the hyperboloid), and pitch of the twist (the smaller the pitch, the more twist there is to the surface).
[*]Root-mean-square difference between the least-squares-fitted strophoid surface and the backbone atoms N, C^α, and C.
[†]Only α-carbon atom coordinates are available for this structure. Consequently, the values obtained are not directly comparable to those obtained for the other structures.

Biochemistry: Novotný and Haber

Proc. Natl. Acad. Sci. USA 82 (1985)   4595

Arg-38 (Lys-38), and Asp-86 in the V_H domain. Figs. 2 and 3 show that (i) conformations of these side chains are identical, within the limits of crystallographic resolution, in all the structures compared; (ii) all the conserved, solvent-exposed hydrophobic side chains are involved in domain–domain association at the bottom of the binding-site barrel and become buried upon formation of the V_L–V_H dimer; and (iii) the buried polar residues are engaged in a conserved hydrogen-bonding network that spans both β-sheets of one domain and tethers such distant parts of the structure as backbone positions 6 and 86. Two hydrogen bonds formed between Gln-38 of V_L and Gln-39 of V_H extend the hydrogen-bonded network across the domain–domain interface and anchor the interface β-sheets in their relative orientation. We propose that all these structural features contribute to invariance of the binding site geometry.

Two-Fold Symmetry of the Binding Site. Figs. 2 and 3 make it apparent that important side chains are related in the V_L–V_H dimer by a pseudo-dyad that is approximately coincident with the axis of the interface β-barrel (21–23). Thanks to this symmetry, Bence Jones proteins (V_L–V_L dimers) are able to associate in the same manner as the V_L–V_H module, creating a domain interface that structurally resembles the antigen-combining site (24, 25) and possesses antigen-binding capacity (26, 27). In fact, cross-sectional dimensions of V_L–V_L interfaces in crystallographic structures REI and MCG correspond closely to those of V_L–V_H domain dimers (Table 1). Side chains at the REI V_L–V_L interface, particularly the pair Gln-38/Gln-38 and the aromatic rings, mimic the side chain arrangement of the KOL V_L–V_H interface (Fig. 4). Solvent accessibility calculations show that surface area buried upon REI V_L–V_L dimerization is smaller by some 3.35 nm^2 than the average V_L–V_H contact area. However, several strong, buried hydrogen bonds provided by residues from hypervariable loops and extending across the V_L–V_L interface supply an additional stabilization in the REI domain dimer (24).

Aromatic Side Chains at the Bottom of the Site. Fig. 2B illustrates the close-packed cluster of the invariant aromatic side chains at the V_L–V_H interface. The clustering is similar to that of other "herringbone" packing motifs (29), characterized by ring centroid distances of approximately 0.56 nm and ring dihedral angles close to 60° (30). Such "perpendicular" ring arrangement is also found in benzene crystals (31). The herringbone geometry principally differs from an apparently directionless packing of aliphatic side chains found at typical β-sheet interfaces (7, 32), and its static and dynamic aspects might be of importance to the process of antigen binding. Numerous experimental data point to small but definite structural rearrangements of antibody molecules upon antigen binding (33–37), and recent crystallographic studies of aromatic ligands bound to the V_L–V_L dimer MCG
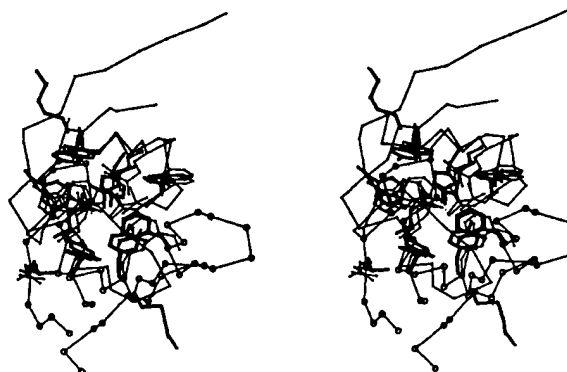
FIG. 4. A close-up of the side-chain arrangement at the V_L–V_L interface of REI. To emphasize the similarity to V_L–V_H interfaces, the backbone segments of KOL that form its binding site are drawn in light lines, together with the prominent side chains that mediate domain–domain contacts between KOL V_L–V_H domains (see also Fig. 2B). The selected domain–domain contacting residues of REI are drawn in heavy lines.

detected rearrangements of aromatic side chains within the binding site β-barrel (38).

Electrostatic Interactions in Variable Domains and Fv Fragments. In computing the electrostatic energy on atoms, residues, and whole domains, we used two different approaches: (i) the model of electrostatics that incorporates an approximate representation of solvent effects (14) and (ii) the unmodified Coulomb formula with the dielectric constant = 50, evaluated to infinity (39). Both methods yielded comparable results and only the solvent-modified energies are reported here. Table 2 shows that the isolated V_L and V_H domains are generally stabilized by electrostatic contributions regardless of their net charge, Σq_i (q_i, the charge of the ith side chain, is $+1$ for lysine and arginine and is -1 for aspartate and glutamate). However, full ionization of acidic side chains is essential for efficient electrostatic stabilization, since the electrostatic energy of structures that were crystallized at lower pH is lower.

The total electrostatic energy of the domains represents a balance between attractive and repulsive side chain interactions. Some of these contributions were found to be very large compared to the resulting total energy; the energy of a single residue, expressed as kJ/residue, may often amount to 20–30% of the total electrostatic energy of the domain (kJ/mol). Residues contributing most significantly are Lys-45, Arg-61, Lys-103, Glu-81, and Asp-82 in the V_L domains and Arg-38, Lys-43, Glu-85, and Asp-86 in the V_H domains. All of them are conserved in other immunoglobulins as well (16) although Arg-38 of V_H is often replaced by a lysine.

Table 2. Electrostatic energy (kJ/mol) and crystallization conditions of immunoglobulin Fv fragments

| Structure | Net charge of domain dimer | Electrostatic potential | | | | V_L–V_H dimer, total | Mother liquor | |
| | | V_L | | V_H | | | | |
		Isolated	In dimer	Isolated	In dimer		pH	$(NH_4)_2 SO_4$, % saturation
KOL	−4	−209	−201	−326	−305	−510	8.0	18
REI*	0	(−180)					8.0	22
MCPC 603	+3	−251	−238	−313	−305	−543	7.0	42
NEW	+4	−58	−125	−125	−155	−263	5.0	42
RHE	−12	−38	+21	−38†	+29†	+54	4.5	26

Crystallization conditions were as described for KOL (46), REI (40), MCPC 603 (41), NEW (42), and RHE (43).
*Since the crystallographic resolution does not permit one to distinguish side-chain amide nitrogens and oxygens in the REI V_L–V_L dimer, its exact electrostatic potential could not be determined. The value given represents an estimate based on arbitrarily assigned amide atoms.
†RHE is a Bence-Jones-type V_L–V_L dimer, not a V_L–V_H heterodimer; the values given in the V_H column refer to the other V_L domain of the V_L–V_L module.
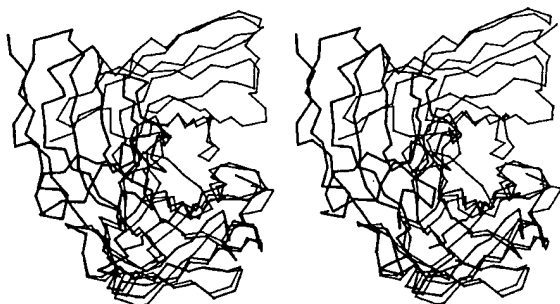
FIG. 5. The mode of V_L-V_L association in RHE. C^α atoms of polypeptide chain backbones are plotted. The four invariant side chains of the first V_L domain of RHE (medium line) were least-squares superimposed on the V_L domain of KOL (light line) as described in the legend to Fig. 1. The second V_L domain of RHE (heavy line) is not matched by this procedure with the KOL V_H domain, as expected if the domain–domain association mode in RHE is comparable to that of V_L-V_H dimers. Rather, the second V_L domain is displaced far to the left of the V_H domain. Electrostatic interaction in the RHE V_L-V_L dimer is repulsive (Table 2), indicating that this dimerization mode might be an artifact of the low pH (4.5) used to crystallize the RHE V_L-V_L dimer.

Lys-45 of V_L and Arg-35 and Lys-48 of V_H belong to those polypeptide chain segments that are directly involved in the binding site. In this sense, the electrostatic stabilization appears to be an indispensable part of the binding site architecture.

Unusual Mode of V_L-V_L Association in RHE. The importance of electrostatic interactions to the integrity of the binding site is next discussed for the structure RHE (44, 45). The two V_L domains of this structure do not dimerize "face to face" as in the V_L-V_H modules but "side by side" (Fig. 5), violating virtually all the characteristics of domain–domain association described above. No close-packed, β-barrel structure exists at the domain–domain interface; instead, the β-hairpin loop of residues 38–48 is displaced some 0.4 nm away from its usual position and makes two interdomain, backbone-to-backbone hydrogen bonds as in regular antiparallel β-sheets. In an apparent correspondence with this anomalous dimerization mode, electrostatic stabilization of the RHE V_L domains is only a fraction of that seen in, e.g., KOL or MCPC 603 domains (Table 2). It would thus appear that a close V_L-V_L association of RHE is only possible under the particular crystallization conditions of extreme hydrogen ion concentration (pH 4.5), where electrostatic interactions are reduced to a small fraction of their original strength and do not significantly enter into the total energetic balance of Gibbs free energy of domain folding and domain–domain association. However, small crystals of RHE were also obtained at pH 6 and their diffraction pattern was reported to be identical to those of the bigger crystals obtained at pH 4.5 (43). Further computational and crystallographic study is needed to clarify the influence of electrostatic force on stability of variable domains and V_L-V_L or V_L-V_H dimers.

1. Marquart, M., Deisenhofer, J. & Huber, R. (1980) *J. Mol. Biol.* 141, 369–391.
2. Saul, F. A., Amzel, L. M. & Poljak, R. J. (1978) *J. Biol. Chem.* 253, 585–597.
3. Segal, D., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. (1974) *Proc. Natl. Acad. Sci. USA* 71, 4298–4302.
4. Richardson, J. S. (1981) *Adv. Protein Chem.* 34, 167–339.
5. Chothia, C. (1984) *Annu. Rev. Biochem.* 53, 537–572.
6. Novotný, J., Bruccoleri, R., Newell, J., Murphy, D., Haber, E. & Karplus, M. (1983) *J. Biol. Chem.* 258, 14433–14437.
7. Chothia, C. & Janin, J. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4146–4150.
8. Chothia, C. & Lesk, A. M. (1982) *J. Mol. Biol.* 160, 309–323.
9. Kabsch, W. (1976) *Acta Crystallogr. Sect. A* 32, 922–923.
10. Padlan, E. A. & Davies, D. (1975) *Proc. Natl. Acad. Sci. USA* 72, 819–823.
11. Amzel, L. M. & Poljak, R. J. (1979) *Annu. Rev. Biochem.* 48, 961–997.
12. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* 112, 535–542.
13. Brooks, B., Bruccoleri, R., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* 4, 187–217.
14. Novotný, J., Bruccoleri, R. & Karplus, M. (1984) *J. Mol. Biol.* 177, 787–818.
15. Northrup, S. H., Pear, M. R., Morgan, J. D., McCammon, J. A. & Karplus, M. (1981) *J. Mol. Biol.* 153, 1087–1109.
16. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. (1983) *Sequences of Proteins of Immunological Interest* (National Institutes of Health, Bethesda, MD).
17. Lesk, A. & Chothia, C. (1982) *J. Mol. Biol.* 160, 325–342.
18. Lee, B. K. & Richards, F. M. (1971) *J. Mol. Biol.* 55, 379–400.
19. Richmond, T. J. & Richards, F. M. (1978) *J. Mol. Biol.* 119, 537–555.
20. Padlan, E. A. (1979) *Mol. Immunol.* 16, 287–296.
21. Davies, D. R., Padlan, E. A. & Segal, D. (1976) *Annu. Rev. Biochem.* 44, 639–667.
22. Padlan, E. A. (1977) *Q. Rev. Biophys.* 10, 35–65.
23. Davies, D. R. & Metzger, H. (1983) *Annu. Rev. Immunol.* 1, 81–117.
24. Epp, O., Colman, P., Feilhammer, H., Bode, W., Schiffer, M. & Huber, R. (1974) *Eur. J. Biochem.* 45, 513–524.
25. Schiffer, M., Girling, R. L., Ely, K. R. & Edmundson, A. B. (1973) *Biochemistry* 12, 1620–1631.
26. Edmundson, A. B., Ely, K. R., Girling, R. L., Abola, E. E., Schiffer, M., Westholm, F. A., Fausch, M. D. & Deutsch, H. F. (1974) *Biochemistry* 13, 3816–3827.
27. Schechter, I., Ziv, E. & Licht, A. (1976) *Biochemistry* 15, 2785–2790.
28. Novotný, J., Bruccoleri, R. & Newell, J. (1984) *J. Mol. Biol.* 177, 567–573.
29. Nockolds, C. E., Kretsinger, R. H., Coffee, C. J. & Bradshaw, R. A. (1972) *Proc. Natl. Acad. Sci. USA* 69, 581–584.
30. Burley, S. K. & Petsko, G. A. (1985) *Science*, in press.
31. Wyckoff, R. W. G. (1969) *Crystal Structures* (Wiley, New York), 2nd Ed., Vol. 6, Part I, pp. 1–2.
32. Cohen, F. E., Sternberg, M. J. E. & Taylor, W. R. (1981) *J. Mol. Biol.* 148, 253–272.
33. Holowka, D. A., Strosberg, A. D., Kimball, J. W., Haber, E. & Cathou, R. E. (1972) *Proc. Natl. Acad. Sci. USA* 69, 3399–3403.
34. Lancet, D. & Pecht, I. (1976) *Proc. Natl. Acad. Sci. USA* 73, 3549–3553.
35. Levison, S. A., Hicks, A. N., Portman, A. J. & Dandliker, W. B. (1975) *Biochemistry* 14, 3778–3786.
36. Schlessinger, J., Steinberg, I. Z., Givol, I. D., Hochman, J. & Pecht, I. (1975) *Proc. Natl. Acad. Sci. USA* 72, 2775–2779.
37. Zidovetzki, R., Blatt, Y. & Pecht, I. (1981) *Biochemistry* 20, 5011–5018.
38. Edmundson, A. B., Ely, K. R. & Hurron, J. N. (1984) *Mol. Immunol.* 21, 561–576.
39. Warshel, A., Russell, S. T. & Churg, A. K. (1984) *Proc. Natl. Acad. Sci. USA* 81, 4785–4789.
40. Palm, W. (1970) *FEBS Lett.* 10, 46–48.
41. Rudikoff, S., Potter, M., Segal, D. M., Padlan, E. A. & Davies, D. R. (1972) *Proc. Natl. Acad. Sci. USA* 69, 3689–3692.
42. Rossi, G., Choi, T. K. & Nisonoff, A. (1969) *Nature (London)* 223, 837–838.
43. Wang, B. C. & Sax, M. (1974) *J. Mol. Biol.* 87, 505–508.
44. Furey, W., Wang, B. C., Yoo, C. S. & Sax, M. (1983) *J. Mol. Biol.* 167, 661–692.
45. Wang, B. C., Yoo, C. S. & Sax, M. (1979) *J. Mol. Biol.* 129, 657–674.
46. Palm, W. (1976) *Hoppe-Seyler's Z. Physiol. Chem.* 357, 799–812.

FUNDAMENTAL IMMUNOLOGY

FOURTH EDITION

Editor

WILLIAM E. PAUL, M.D.

Laboratory of Immunology
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, Maryland

Printed in the United States of America

9 8 7 6 5 4 3 2 1

CHAPTER 3

Immunoglobulins: Structure and Function

J. Kimble Frazer and J. Donald Capra

INTRODUCTION

Immunoglobulin is the crux of the humoral immune response. As a cell surface receptor on B lymphocytes, immunoglobulin is responsible for instigating cellular processes as diverse as activation, differentiation, and even programmed cell death. As secreted antibody in plasma and other bodily fluids, immunoglobulin is able to bind foreign antigen, thereby either neutralizing it directly or initiating steps necessary to arm and recruit effector systems such as complement or antibody-dependent cell cytolysis by monocytic phagocytes. The ability of immunoglobulin to perform such a wide array of duties can be attributed to evolution's clever usage of a structural paradigm—the immunoglobulin domain—and its duplication, diversification, and elaboration upon that design to endow it with an assortment of functional qualities.

Despite the variety of purposes served by immunoglobulin molecules, one feature remains common to virtually all considerations of immunoglobulin structure and function: immunoglobulins have an amazing capacity to interact with other molecules. In one sense, immunoglobulins must be able to effectively bind a finite set of invariant partners, such as Fc receptors, signal-transducing molecules, and components of the complement cascade. In another sense, immunoglobulins, collectively, must meet the challenge of

J. K. Frazer: Graduate Program in Molecular Microbiology and Medical Scientist Training Program, University of Texas Southwestern Medical School, Dallas, TX 75235-9140.
J. D. Capra: President and Scientific Director, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104.

being able to recognize an essentially infinite array of antigenic determinants. More remarkable, perhaps, is the fact that immunoglobulin is frequently called upon to fulfill both of these binding responsibilities simultaneously, and in such a way as to mediate significant biological effects. As such, immunoglobulin molecules may be viewed as a marriage between the constraints engendered by biological continuity and the quest for diversity superimposed upon this evolutionary framework.

The lengths to which evolution has gone in order to bestow immunoglobulin with these conflicting capabilities has been the subject of intense scientific scrutiny, and has yielded innumerable fascinating insights into immunology, genetics, protein chemistry, and the discipline of biology as a whole. In trying to understand how antibody is able to recognize such a multitude of different specificities, science has benefited from the discovery of both VDJ recombination (see Chapter 5) and somatic hypermutation (see Chapter 25). In an attempt to reconcile the incongruity entailed by the observation of highly divergent N-terminal regions coupled to constant C-terminal domains, research has gained not only the once-heretical "two genes, one polypeptide" hypothesis (1), but also the concept of isotype switching (see Chapter 24). Thus, studies into immunoglobulin diversity have proven to be extremely profitable scientific endeavors. In addition, while diversity has been a hallmark of the study of immunoglobulin since it was first recognized to be a salient feature, several aspects that derive from immunoglobulins' underlying uniformity have been used to glean understanding into protein structure–function relationships in general.

Immunoglobulins were the first molecules described from the ancestral immunoglobulin superfamily (IgSF) (2–4). As an ever-

expanding gene family, members of the IgSF have been shown to be vital to issues of cell–cell interaction and molecular recognition in a variety of cell types and across several taxonomic boundaries. Many molecules central to the functioning of the immune system, including the antigen-specific chains of the T-cell receptor (TCR) (see Chapter 10) and the class I and II major histocompatibility complex (MHC) antigens (see Chapter 8), are counted among this group. Common to all members of the superfamily is the presence of one or more immunoglobulin-like domains. Three-dimensional structural analyses of proteins containing these regions have demonstrated that the conserved amino acid sequences that make up an immunoglobulin homology domain comprise a recurring structural motif that can fold into a compact globular subunit. These subunits, in turn, are capable of integrating into complex macromolecules (5,6). As a result, different immunoglobulin molecular structures are similar not only to each other, but also to a multitude of other important proteins.

Because such a concerted scientific effort has been made to understand the way in which immunoglobulin functions, a large volume of sequence information—at both the nucleotide and amino acid levels—is available in both the literature and public databases. Indeed, immunoglobulins have likely been sequenced more frequently than any other class of gene or protein. Similarly, immunoglobulins have been well represented in structural studies, crystallographic and otherwise, to an unprecedented degree. This mass of work has allowed a number of conclusions regarding structure–function relationships of immunoglobulins to be made. Specifically, the aim of this chapter is not to compile an exhaustive catalogue of all extant work on the topic of immunoglobulin sequences, but rather to present the essential features of immunoglobulin structure and their relation to immunologic function as is currently understood. Further, because immunoglobulin proteins have been so evolutionarily valuable, they can be found, in one form or another, throughout vertebrate species. Many of these molecules are only now being characterized, and surely many more are yet to be identified. As a consequence of this diversity, however, it is impossible to relate all of the details of immunoglobulin structure and function in their entirety. Instead, unless otherwise noted, the examples of human and murine immunoglobulins will be used as models to convey the general conclusions garnered from scientific insight and experimentation into this important and fascinating class of proteins. The organization of this discussion will begin, following an introduction to basic immunoglobulin features, with a consideration of the primary structure of antibody molecules and proceed through the secondary, tertiary, quaternary, and higher order immunoglobulin structural topics that derive from its sequence. Once this foundation has been laid, the functional attributes of immunoglobulin will be considered, with an eye to correlating an antibody's capacities—to the extent which it is possible—with that of its structure. A section on the IgSF follows, which will briefly address its evolution and also specifically detail particular IgSF members critical to immune responses that are not explicitly covered elsewhere in this volume.

GENERAL IMMUNOGLOBULIN STRUCTURE, NOMENCLATURE, AND HISTORY

Structural Considerations

Figure 1 presents a diagrammatic representation of an antibody molecule. The typical immunoglobulin monomer is comprised of
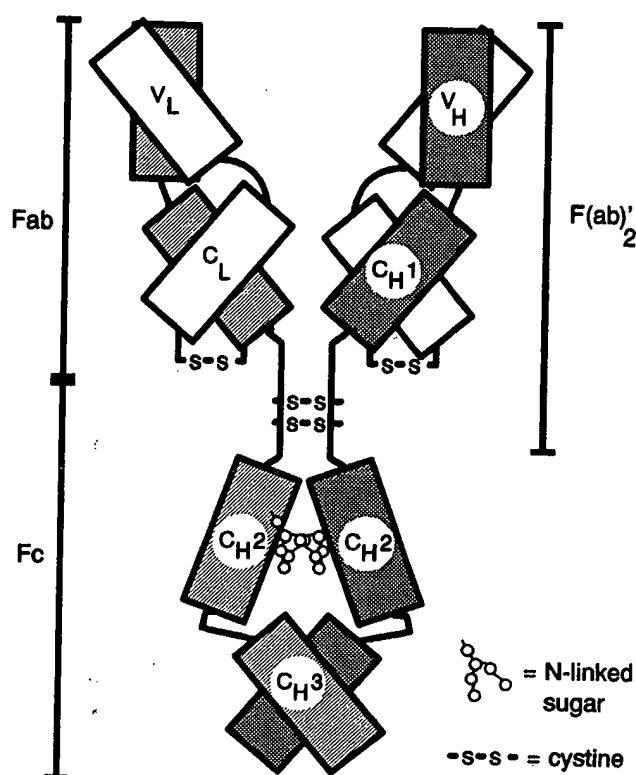
FIG. 1. Schematic representation of a prototypic immunoglobulin monomer. Each box symbolizes a complete immunoglobulin domain from either the heavy (shaded boxes) or light (unshaded) chain. Labeling of domains follows standard nomenclature, as outlined in the text. Interchain disulfide bonds are denoted by black bars. Note that these bonds are present between both heavy and light chain pairs and between the two heavy chains. Conserved N-linked carbohydrate occurs on all C_H2 domains as shown, although some immunoglobulins are also glycosylated at additional sites elsewhere in the molecule. Also of note is the fact that all of the domains associate to form dimeric modules (V_H/V_L, C_H1/C_L, and C_H3/C_H3), except C_H2 domains. The Fab, Fc, and F(ab)'$_2$ proteolytic fragments are demarcated by bars to either side of the diagram. (From ref. 6a, with permission.)

four polypeptide chains complexed together via hydrophobic interactions and stabilized by disulfide bonds. Due to allelic exclusion (see Chapters 5 and 6) B lymphocytes usually express only one functionally rearranged heavy chain gene and only one light chain polypeptide as well. Consequently, complete immunoglobulin proteins are composed of two identical heavy chain polypeptides of approximately 55 kD and two identical light chains of 25 kD. Each heavy and light chain pair is joined by one or more interchain disulfide bonds, and also relies upon non-covalent interactions to properly orient the two chains relative to each other. One such "half-antibody" contains a single antigen binding site (i.e., it is monovalent). The complete four polypeptide chain monomer is formed by similar hydrophobic bonding between the two heavy chains, and it also utilizes one or more disulfide bonds to stabilize the complex. Thus, a complete immunoglobulin molecule is bivalent with two identical sites for potential binding of antigen. As such, an immunoglobulin may be thought of as a "dimer of a heterodimer," although these half-molecules do not occur naturally.

Each individual polypeptide chain consists of two to five domains of approximately 110 amino acids (7), each capable of folding independently. These domains form compact, protease-resistant structures which serve as the fundamental unit of immunoglobulin structure. The interactions that allow for the formation of the aforementioned immunoglobulin monomer almost exclusively occur in pair-wise fashion between domains of two different polypeptide chains (see Fig. 1), such that the functional modules of an antibody are in fact dimerized domains. In addition, as each domain of an antibody molecule is encoded by a separate exon, immunoglobulin domains also serve as the essential element of antibody genetics. In this light, it is easy to recognize how evolution has used the prototypical immunoglobulin domain as a substrate for experimentation, and as a result different domains have attained distinct structural and functional attributes. Moreover, the presence of one or more "immunoglobulin homology domains" also proves to be the distinguishing characteristic for inclusion in the immunoglobulin gene superfamily. Thus, the duplication and adaptation of the Ig homology domain has occurred not only within the context of formal "immunoglobulin genes," but also in the greater scope of the IgSF, which far predates the emergence of antibody. In either case, the archetypal immunoglobulin domain has clearly proven to be a powerful evolutionary tool, as will be detailed below and in greater detail throughout this chapter.

The hallmark of all Ig domains is the presence of a structural motif termed the *immunoglobulin fold*. This characteristic feature is actually a specialized "β-barrel" typically comprised of seven polypeptide strands, which form antiparallel β-pleated sheets in the folded domain. This configuration is depicted in Fig. 2, which was deduced from x-ray diffraction studies of an immunoglobulin light chain (8). Each Ig domain is composed of two β-pleated sheets, one containing four β strands, the other consisting of at least three β strands (represented by arrows in Fig. 2). Loops of variable length connect the different strands, allowing the β sheets to form. The two β-pleated layers are oriented in a sandwich, enclosing a hydrophobic interior. Further stability is provided by a disulfide bond near the domain's core, which covalently links the two sheet layers. The cysteines that contribute this bond are conserved in all immunoglobulins, and in almost all proteins that possess Ig-like domains. Two residues, a tryptophan in strand 3-1 and an aromatic residue that precedes the second half-cystine, are also maintained consistently and serve to protect the disulfide bond in the three-dimensional structure. Other conserved features include hydrophobic core residues, which stabilize the inside of the sandwich, and glycine and proline loop residues, which provide the flexibility necessary for the formation of these interconnecting sequences (9–12).

Since the hydrophobic core residues are predominantly responsible for promoting the folding of the β sheets, and thus the entire immunoglobulin fold, the sequences of the loop residues are free to vary considerably. This, in turn, grants loop residues the freedom to serve as substrates for selection, at the level of selection of a particular antibody in an immune response and at the level of natural selection in phylogeny. In this way, the prototypical immunoglobulin homology domain serves as a potent cofactor for the evolution of both organismal immunity and that of the species in general.

Immunoglobulin Nomenclature

Light chains contain two such immunoglobulin domains, whereas a heavy chain is made up of either four or five domains, depending on the type of heavy chain (isotype) used by the antibody in question. Different immunoglobulin domains possess different structural and functional characteristics, and their naming, in part, reflects these differences. The amino-terminal domain of each chain, whether of the heavy or light type, is termed a *variable* (V) region due to the discovery of extensive sequence divergence between different antibody proteins in this part of the molecule. These are designated V_H and V_L for heavy and light chains, respectively. V regions have been demonstrated to be responsible for the antigenic specificity of the immunoglobulin.

Carboxy-terminal domains, on the other hand, display considerably less sequence variation within a given isotype and are referred to as *constant* (C) regions. Heavy chain C regions are numbered C_H1, C_H2, and so on, beginning with the most V region-proximal domain. The constant region domains of the heavy chain have been shown to be responsible for many aspects of antibody function, including interaction with Fc receptors, complement fixation, transplacental transfer, the ability to multimerize, and the capacity to be secreted on mucosal surfaces. Because different heavy chain isotypes have different C region domains (i.e., the C_H3 domains of different isotypes are distinct), these capabilities vary with the class of the particular antibody. Five major classes of heavy chain C regions exist: alpha (α), gamma (γ), delta (δ), epsilon (ε), and mu (μ). As a direct consequence of the correlation between the
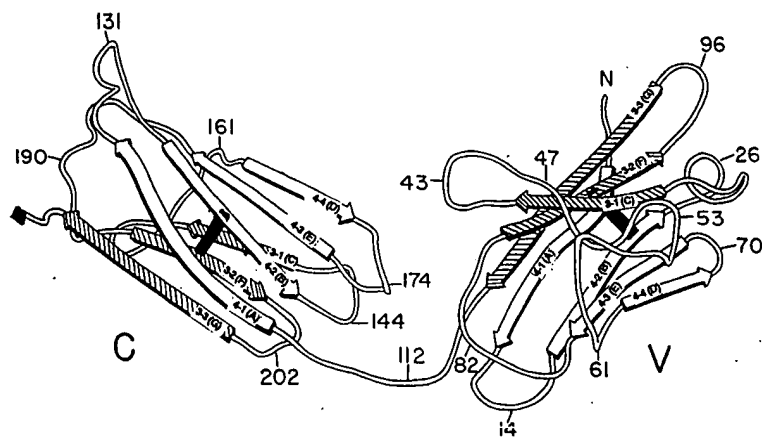
FIG. 2. Ribbon drawing of the V and C domains of a light chain. β strands are depicted as *arrows,* with those of the four-stranded face *unshaded* and those of the three-stranded face *shaded.* Strands are numbered according to Edmundson and lettered (*in parentheses*) according to Hood. Intrachain disulfide bonds are represented as *black bars.* Selected amino acids are numbered, with position 1 being the N-terminus. Residues 26, 53, and 96 correspond to amino acids in CDRs 1, 2, and 3, respectively. The dimerization surfaces of each domain (four-strand side of the C domain, three-strand side of the V domain) face upwards. (Adapted from ref 8, with permission.)

heavy chain class of an antibody and its resultant effector functions, immunoglobulins are named according to their heavy chain, using an English-letter terminology (IgA, IgG, IgD, IgE, IgM), which corresponds to their Greek letter isotypes. Specific domains of C regions are often designated according to the class of heavy chain from whence they originate as well (i.e., the C_H3 domain of a μ antibody is signified by $C_\mu3$). Owing to the propensity of immunoglobulin domains to evolve independent of one another, oftentimes a particular domain of a specific isotype may be responsible for one or more functional characteristics of the entire antibody, making this naming system particularly relevant. On the other hand, the constant regions of light chains, possessing only one C domain, are usually simply denoted by C_L. The two light chain classes, kappa (κ) and lambda (λ), may be indicated by the use of C_κ or C_λ designations. No distinct functional attributes have as yet been ascribed to either the κ or λ light chain isotypes.

Finally, immunoglobulins also have hinge regions located C-terminal to the C_H1 domains of their heavy chains. In the case of heavy chains of the μ and ϵ isotypes, the hinge is so elongated that it is actually an extra immunoglobulin domain, explaining the presence of a fifth C domain in these molecules. Other heavy chain classes use shorter stretches of protein, which are thought nonetheless to have evolved from the $C_{\mu/\epsilon}2$ domain. Consistent with the independent evolution of the other domains of immunoglobulin genes, hinge regions are generally encoded by individual exons as well. As the name implies, the hinge permits a generous degree of flexibility between the antigen-binding and effector-interacting components of the molecule. Thus, the hinge domain facilitates linking of the two disparate elements of immunoglobulin function: the ability to interact with an endless array of structural determinants on antigenic surfaces (mediated by V regions) and the capacity to interact with a limited number of effector-activating molecules (mediated by C regions). In addition, disulfide bridges between the two heavy chains typically occur within the boundaries of the hinge region, allowing the complete tetrameric complex to form.

Hence, in many cases, the discrete elements of immunoglobulin structure defined both genetically and structurally as immunoglobulin domains are also responsible for specific functional qualities. Moreover, in addition to the one-domain-per-exon correlation that exists for immunoglobulins, in both heavy and light chains the V region domain and the domains of the different C regions are in fact distinct genes (13). This type of genetic arrangement allows the ability to recognize a specific antigen to be united with the effector functions that are most appropriate for that particular immune response at that particular time. In this regard, then, it is clear that antibodies truly embody the linkage of structure to that of function.

An Historical Perspective

Long before x-ray diffraction of crystals had yielded the keys to dissecting the structure of immunoglobulin, other seminal studies had been performed that, in retrospect, agree completely with the conclusions drawn from crystallization analysis. Many of these experiments focused on the basic protein chemistry of pooled IgG, using the techniques of proteolysis, reduction, and denaturation. First, it was revealed that papain digestion of IgG would render two types of protein fragments: Fab, a monovalent antigen binding fragment, and Fc, an easily crystallizable fragment (14). Soon after, it was recognized that pepsin treatment of IgG produced an antigen-binding fragment designated F(ab)'$_2$ which had bivalent activity (15). Furthermore, if this fragment was treated with reducing agents, two univalent Fab' fragments could be obtained. These different fragments are schematically represented in Fig. 1. Reduction and dissociation of IgG also demonstrated that identical heavy and light chains were complexed via disulfide bonds (16). This and other work eventually allowed investigators to decipher a working model for immunoglobulin structure consisting of four polypeptide subunits—two identical heavy chains and two identical light chains—stabilized by multiple interchain disulfide bonds (17,18), which we now know to be correct.

Early studies of a different type also proved successful in revealing information about antibodies; these experiments focused upon utilizing the immune system itself as a means to decode aspects of immunoglobulin structure (reviewed in refs. 19–21). As large glycoproteins, antibodies themselves are potent immunogens capable of eliciting vigorous humoral immune responses. Investigators used immunoglobulin preparations (initially either heterogeneous total serum immunoglobulin or homogeneous myeloma or plasmacytoma proteins, and later monoclonal antibodies from hybridomas) as antigens to generate antibody responses by injecting them into animals of differing species or different animals of the same species. The antibodies produced by these immunization protocols proved useful in resolving several key elements of immunoglobulin structure, and many of the antigenic determinants recognized by these antisera have subsequently been shown to correlate exactly with known structural features of immunoglobulins.

A three-tiered serological classification scheme for immunoglobulin was devised using these antisera (after adsorption) as reagents to categorize antibody molecules into distinct groups. The first tier of organization is that of the *isotype*. Isotypes define C region determinants and, as such, distinguish heavy and light chain constant regions from one another. Initially, five heavy chain classes were recognized and given the Greek letter designations mentioned in the previous section. The presence of these five isotypes in virtually all mammals for which immunoglobulin profiles have been determined indicates that the divergence of C region genes occurred at an early stage of mammalian evolution. Similarly, light chain constant regions were also divided into discrete κ and λ classes. Soon after, refinements made clear that two of the human heavy chain classes, α and γ, in fact contained several related members that could be further divided into subclasses. Human IgA is separated into $\alpha1$ and $\alpha2$ subclasses, and human IgG is separated into four γ subclasses: $\gamma1$, $\gamma2$, $\gamma3$, and $\gamma4$. Murine IgG is also composed of four γ subclasses ($\gamma1$, $\gamma2a$, $\gamma2b$, and $\gamma3$), although their structural and functional characteristics—and their abbreviated designations—do not agree with their human counterparts, thus indicating this diversification occurred after these species' evolutionary divergence. Each different isotype (whether class or subclass) is represented by a separate C region gene in the haploid genome, and all isotypes are present in the sera of all normal individuals of a given species.

Allotypes, on the other hand, refer to determinants found on the antibodies of some, but not all, members of a species. These determinants are encoded by one allele of a particular C region gene (either heavy or light chain) and are inherited in typical Mendelian fashion as autosomal dominant traits. Compilations of human allotypes are summarized in ref. 22 and are covered more extensively in ref. 23. Whereas isotypes and allotypes are localized to the C regions of immunoglobulins, *idiotypes* are antigenic determinants found on the V regions of antibodies, and they frequently correlate

with binding specificity. Generally, idiotypes are present only in an individual member of a given species, and these antigenic epitopes reflect the uniqueness of each individual immunoglobulin molecule. An idiotypic determinant defined by a monoclonal antibody is called an *idiotope*. Idiotypes are not always restricted to the individual, however. Occasionally, when two individuals are challenged with the same antigen, they will produce antibodies that share the same idiotypic determinant(s). In such cases, the idiotype is called a *cross-reactive*, or *public*, *idiotype* (24). Cross-reactive idiotypes represent the usage of the same V gene segment by different individuals. Thus, idiotypes may be best thought of as being restricted not to the individual organism, but to the individual immunoglobulin molecule.

Obviously, a tremendous amount of effort, using a variety of scientific approaches, has been focused upon attempts to understand immunoglobulin structure. It is remarkable, though, that the vast majority of this early work, whether utilizing protein chemistry to resolve basic structural characteristics or manipulating the humoral immune response to generate reagents to classify immunoglobulin proteins relative to one another, has in fact identified, and in many cases answered, many of the crucial questions of immunoglobulin structure correctly. As shall be discussed in the following sections, many of the crucial structural features of the antibody—from primary sequence to quaternary associations—were first inferred from these initial landmark studies.

IMMUNOGLOBULIN STRUCTURE

Primary Structure—Two Genes, One Polypeptide

The assertion that each immunoglobulin chain derived from two distinct genetic entities was a novel and provocative hypothesis at the time it was proposed (1), and it has proven to be correct. Owing to this genetic independence, the structures of the V and C regions of immunoglobulin will be treated separately here as well, although they obviously share numerous commonalties.

Among the most remarkable discrepancies between V and C regions are the differences in their genetic organization. While the different heavy and light chain C regions are encoded by fewer than 20 genes, V region genes (VH, Vκ, and Vλ) number in the hundreds. Further distinguishing the V region loci is the fact that genes for complete VH or VL domains are not present in the genome originally, but are "recombined" at the genetic level according to the processes of somatic diversification described in Chapter 5. This recombination of V, D, and J elements to form functional heavy chain genes (or V and J in the case of light chains) imposes another degree of diversity upon the V regions. Due to these complexities, sequence variability is in fact the hallmark of variable region domains. In addition to the differences between V and C domains in their genetic design and construction, V region sequences also have uniquely identifiable characteristics. One of the most easily recognizable features is the fact that V regions are approximately 16 residues longer than the prototypic 110 amino acid immunoglobulin domain. These extra residues allow V regions to form a distinctive immunoglobulin fold structure using two additional β strands, which distinguishes V domains from C domains and also has implications for V region function. Also, the processes of V(D)J recombination can further alter the germline—encoded length of V regions, subtly affecting V domain structure and function as well.

V regions form the amino-terminal domains of heavy and light chains, and their primary responsibility is the binding of antigen.

The promotion of the capacity to recognize antigenic determinants has been the driving force behind the evolution of V genes, both at the structural level of individual genes and for the evolution of the different V loci. When sequence data first became available from antibody proteins (see Fig. 3), it was apparent that great variation existed between V regions relative to that found between C regions. A means was developed to quantitate this variation whereby variability was defined as the number of different amino acids observed at a given position divided by the frequency of the most common amino acid at that position (25). Using this equation, an invariant residue would have a variability equal to one, whereas the theoretical upper limit for a position occupied by each of the 20 amino acids in a random fashion would be 400. This can be illustrated graphically by plotting the variability scores of a particular protein against its residue number, as is demonstrated in Fig. 4. Variability plots of this type established not only that V regions were characterized by diversity in their sequences, but that this variation was principally clustered in three regions, which were deemed *hypervariable regions* (HVRs). It was hypothesized that these highly variable segments of heavy and light chains would coordinate in such a way as to form the antigen-combining site

```
LIGHT CHAIN VARIABLE REGIONS

        1                              A 27 F
AG:  DIQMTQSPSSLSASVGDRVTITCQASQ------DINHYL
LEN: **V*****NS*AV*L*E*A**N*KS**SVLYSSNSKN**
TI:  E*VL***GT**L*P*E*A*LS*R***S-----VSNSF*


AG:  NWYQQGPKKAPKILIYDASNLETGVPSRFSGSGFGTDFT
LEN: A****K*GQP**L***W**TR*S***D******S*****
TI:  A****K*GQ**RL***V**SRA**I*D******S*****


                                            109
AG:  FTISGLQPEDIATYYCQQYDTLPRTFGQGTKLEIKRT
LEN: L***S**A**V*V******YST*YS**********T
TI:  L***R*E***F*V******GSS*S*******V*L**T


KAPPA LIGHT CHAIN CONSTANT REGIONS

     108
AG:  TVAAPSVFIFPPSNEQLKSGTASVVCLLNNFYPREA
LEN: ************************************
TI:  ************D***********************


AG:  KVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLT
LEN: ************************************
TI:  ************************************


                                       214
AG:  LSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC
LEN: ************************************
TI:  ************************************
```

FIG. 3. Amino acid sequences of human antibody light chains. *Dashes* denote gaps introduced to optimally align sequences; *asterisks* represent identity relative to the top sequence. These are among the first immunoglobulin sequences ever obtained, demonstrating the dramatic differences in variation between V and C regions. The appearance of clusters of conserved residues within the different V domains also illustrates the necessity of a system to accurately quantify variations between several sequences (see Fig. 4). (Developed from the sequence compendium of Kabat et al. in ref. 24a.)

HUMAN HEAVY CHAINS

FIG. 4. Variability plot of human heavy chains. The hypervariable regions are apparent as the three obvious peaks in the graph. (From ref. 24a, with permission).

(reviewed in ref. 26); thus they were termed *complementarity-determining regions* (CDRs) as well.

Variability analysis also determined that other stretches of V region sequence were reasonably well conserved from protein to protein; these were presumed to perform basic structural functions necessary for proper folding of all V domains. Accordingly, they were dubbed *framework regions* (FRs) because they provide the platform that supports the CDRs. Structural analyses have confirmed that the FRs largely coincide with the β strands of the immunoglobulin fold, while CDRs, on the other hand, chiefly correspond to the loops that join β strands on the C region-distal end of the V domain. A linear representation of this association is shown in Fig. 5; note that CDRs 1, 2, and 3 join β strands B and C, C and D, and F and G, respectively. Significant sequence motifs are also apparent in this comparison of six V region proteins: a W/F-G-X-G motif in FR4 that is common to all V domains (27), the V_H-specific G-L-E-W-hydrophobic stretch in FR2 (28), and the V_L-specific sequence P-hydrophilic-hydrophobic-L-hydrophobic in the analogous FR2 location (28). These motifs are vital for proper dimerization of domains and will be discussed further in the section on quaternary immunoglobulin structure. Another important distinction between heavy and light chain V region sequences is also apparent in Fig. 5: relative to V_L domains, V_H regions generally utilize longer FR1 and CDR2 segments and shorter CDR1 and FR2 stretches. While the vast number of V genes precludes the ability to definitively assign boundaries for FRs and CDRs that are constant among all immunoglobulin V regions, Table 1 summarizes the traditional positions that delineate these areas for both heavy and light chains.

The presence of hundreds of different germline V region genes obviously contributes greatly to the sequence diversity of different variable domains. However, the somatic process of V(D)J recombination (see Chapter 5) further accentuates V region variability, specifically targeting the CDR3 of the protein (29). In this system, approximately 100 unique V genes (V_H, V_κ, or V_λ loci) encode the

N-terminal FR1-CDR1-FR2-CDR2-FR3-5'CDR3 portions of V regions, while four to six "joining" (J) minigenes code for the carboxy-terminal 3'CDR3-FR4 segments. Heavy chains also incorporate one of about 30 short "diversity" (D) gene segments between V and J genes to generate complete V region domains. The relationship between rearranged V(D)J gene segments and the FR/CDR organization of the V region is schematically represented in Fig. 6. The combinatorial assortment of gene segments to form complete heavy and light chains, followed by the combinatorial assortment of heavy and light chains with each other to form antigen-binding V_H/V_L dimers, results in a practically limitless number of V domain structures. Moreover, during the recombination process itself, the activity of exonucleases and untemplated N-segment additions (29), templated P-nucleotide incorporation (30), and D-D fusion events (31) can boost the diversity of CDR3 even further. Finally, superimposed on these aspects of "combinatorial" and "junctional" diversity-generating mechanisms, somatic hypermutation (32,33; see Chapter 25) serves to introduce still more variation by altering residues throughout the V region.

Despite—and perhaps as a result of—the seemingly endless number of possible V region sequences, sophisticated schema have emerged for their classification. These groupings are based upon homology-based hierarchies that directly reflect the evolution of the antibody gene loci. Members of a group are more similar to each other than to all other sequences from other groups and share linked amino acid substitution patterns, which serve as "identifiers" for the various classifications. The most evolutionary distant grouping is, of course, that of the V regions themselves, followed by the V_H, V_κ, and V_λ distinctions, which represent the separate V gene loci. In humans, the heavy chain locus is found on chromosome 14, and the κ and λ loci are found on chromosomes 2 and 22, respectively (34–36). In the mouse, these genes are located on chromosomes 12, 6, and 16 (37–39). Other stratifications for V region organization also mirror the evolution of the antibody gene loci. The use of "clans" to categorize V genes has demonstrated the

4-1 (A)	4-2 (B)	3-1 (C)

```
V_λ1 (WAH)  Q S V L T Q P P S A S G T P  Q R V T I S  C  F G S S S N I G R Y - Y V Y  W Y Q  L P  Q  T T P K L L I Y · · · · · K D N
V_λ2 (BUR)  Q S A L T Q P R S V S G S P  H S V T I S  C  I G T S S N V G D Y K Y V S  W Y Q  H P  G  K A P K L I I Y · · · · · E V S
V_κ1 (CAR)  D I Q M T Q S P S T L S A S V  D R V A I T  C  R A S Q · · · N I S S W L A  W Y Q  K  K  K A P K V L I Y · · · · · K S S
V_κ3 (B6)   D I V L T Z S P G T L S L S P  Z R A A L S  C  R A S Q S · · L S G N Y L A  W Y Q  K P  Q  Q A P R L L M Y · · · · · G V S
V_H1 (EU)   Q V Q L V Q S G A E V - K K P  G S S V K V S  C  K A S G G T F S R S A - I I  W V R Q A P  G  Q G L E W M G G I V P M F G P P
V_H3 (ZAP)  E V Q L V E S G G A L - V Q P  G G S G R L S  C  A A S G F T F S T T S - R F  W V R Q A P  G  K G L E W V E F R V Q G S A I S
```

	FR1	CDR1	FR2	CDR2
V_λ	FR1	CDR1	FR2	CDR2
V_κ	FR1	CDR1	FR2	CDR2
V_H	FR1	CDR1	FR2	CDR2

4-4 (D)	4-3 (E)	3-2 (F)	3-3 (G)

```
V_λ1  Q R P S G V P D  R - - F S G S K  S  G T S A S  A I S G L R S E  D E A D Y Y C  A A W D D S L W V F G G G T T L T V L S
V_λ2  S R P S G V P D  R - - F S G S K  A  G N T A S  T I S G L Q A E  D E A D Y Y C  C S Y I G S - Y V F G T G T K V I V L G
V_κ1  S L E S G V P S  R - - F S G S G  S  G T D F T  T I S S L Z P B  D F A T Y Y C  Q Q Y N - T F F T F G P G T K V D I K R
V_κ3  S R A T G I P D  R - - F S G S G  S  G A D F T  T I S R L Z P E  D F A V Y Y C  Q Q Y G S S P F T F G Q G S K L E I K
V_H1  N Y A Q K F Q G  R  V T I T A D E  S  T N T A Y  M E L S S L R S E  D T A F Y F C  A G G Y - G I Y S P E E Y N G G L V T V S S
V_H3  H Y A D S V Q A  R  F T I S R N D  S  K N T L Y  L Q M N T G E A Z D  T A V Y Y C  A R T R P G G Y F S D V W G Q G L V T V S S
```

	CDR2	FR3	CDR3	FR4
V_λ	CDR2	FR3	CDR3	FR4
V_κ	CDR2	FR3	CDR3	FR4
V_H	CDR2	FR3	CDR3	FR4

FIG. 5. Sequence alignment of six human V regions. Boxes above sequences represent β strands of the domains. Strands are numbered according to Edmundson and lettered (in parentheses) according to Hood. Gaps introduced to maximize homology are represented by dashes. Amino acids conserved among all six proteins are boxed and shaded. Boxes beneath the sequences depict the statistical boundaries of V region subdomains (see Table 1). Note differences in the lengths of the FR1, CDR1, FR2, and CDR2 segments between the four light chains and two heavy chains.

TABLE 1. Boundaries delineating the statistical and structural subdomains of variable regions

Subdomain region	Ig Chain	Residue positions	Boundaries of structural loop
FR1	Heavy	1–30	
	Light	1–23	
CDR1	Heavy	31–35*	H1: 26–32
	Light	24–34*	L1: 26–33
FR2	Heavy	36–49	
	Light	35–49	
CDR2	Heavy	50–65*	H2: 53–55
	Light	50–56	L2: 50–52
FR3	Heavy	66–94*	
	Light	57–88	
CDR3	Heavy	95–102*	H3: 96–101
	Light	89–97*	L3: 91–96
FR4	Heavy	103–113	
	Light	98–107*	

FR1–4 (framework regions), CDR1-3 (complementarity-determining regions), H1-3 (heavy chain variable loops), and L1-3 (light chain variable loops) are numbered according to Kabat et al. (1) using an alignment giving priority to conserved residues. Asterisks indicate regions which may have length variations depending upon germline V gene usage and/or junctional diversity. Data for the table were compiled from Kabat et al. (24a) and Chothia and Lesk (28a).

A.

B.
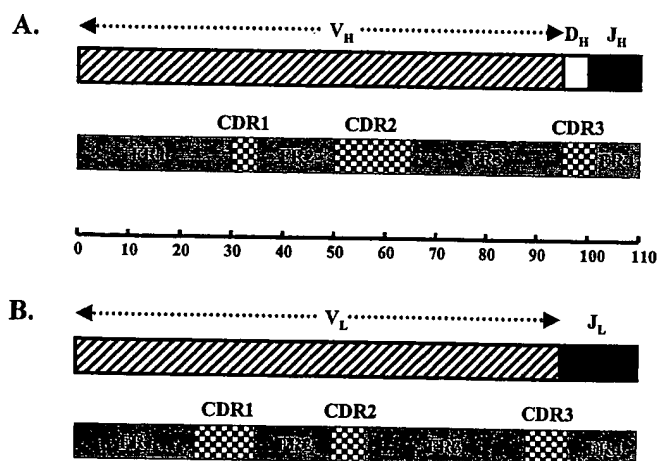
FIG. 6. Comparison of the gene structures and protein subdomains of rearranged V regions. (A) Heavy chain V domain. (B) Light chain V domain. The V_H and V_L gene segments are represented as hatched boxes, the D_H gene segment as a white box, and the J_H and J_L gene segments as black boxes. Framework regions (FRs) are displayed as shaded boxes and complementarity-determining regions (CDRs) are pictured as checkered boxes. An approximate amino acid scale separates the two diagrams.

development of V loci across several vertebrate species (see Fig. 7) for both heavy and light chains (40–44). Three V$_H$ clans have been recognized using nucleotide sequence homology comparisons across the FR1 6–24 codon interval. While this stretch of FR1 sequence is conserved within a clan, a similar span (the 67–85 codon interval of FR3) can also be used to discriminate between V$_H$ genes that belong to the same clan but differ in regard to the next level of classification, that of the family.

Classically, families are the groupings that have been used most frequently to identify and categorize V region genes relative to one another (reviewed in refs. 45–47). Members of a V region family share about 80% identity at the DNA level. Historically, when genomic Southern blotting was used to work out approximate family sizes, this degree of homology allowed for sufficient cross-hybridization to occur under low-stringency conditions, accounting for the utility of the family designation. At the protein level, this translates to levels of about 75% identity between gene products from the same family and less than 70% homology for proteins belonging to different families. Using these criteria, human V$_H$ genes may be segregated into seven families, V$_κ$ into four major families, and V$_λ$ into ten families. The murine system (which contains larger absolute numbers of V genes) is more complicated, as evidenced by the fact that 14 V$_H$ and 20 V$_κ$ families have been recognized.*Example sequences from several human V$_κ$ and V$_H$ gene families are aligned in Fig. 8. Note the presence of numerous "shared substitutions" within pairs of sequences belonging to the same family that are not present in the other sequences. These serve as distinctive "signature residues," which facilitate rapid identification of a particular sequence's "family of origin." The sequences in Fig. 8 also demonstrate another important characteristic of V gene families: Different families frequently possess different CDR lengths. Thus, independent of amino acid sequence, V

region families intrinsically possess differing binding-site structures, thereby affecting their functional capabilities.

Shared substitutions have also been used to further refine families into subfamilies. Subfamilies, as are all classification schema to some extent, are particularly arbitrary divisions, largely because the parameters used to define subfamilies are not standardized. Generally, the features that describe a particular subfamily are esoteric and depend upon the specific characteristic(s) being studied. Finally, at the most descriptive level of classification, the work of Rabbitts, Winter, Honjo, and many others (46–50) has resulted in the identification, mapping, and—to some extent—sequencing of presumably most, if not all, human V region genes. Within a family, single V gene segments may be compared against a consensus sequence representing that family (Fig. 9). When such a comparison is performed, it becomes clear that individual V genes' divergence is focused primarily in their CDR1 and CDR2 segments (42). Thus, on the basis of primary sequence structural information, even closely related V genes may be predicted to adopt similar framework cores but differ in terms of their CDR1 and CDR2 loops such that a plethora of potential antigenic specificities are encoded genomically. In fact, even at the level of the individual V gene variability persists in this system. As one might expect with "variable" genes, allelic variation and polymorphism of the antibody loci exists as well—such that it is probably safe to conclude that all of the possible incarnations of immunoglobulin V genes will never truly be identified and categorized. Suffice to say, then, that the variable gene loci, clans, families, subfamilies, and even individual V genes themselves all derive from gene duplication events; only the period in evolutionary time at which the duplication occurred truly separates one gene or group of genes from any another.

While V regions provide the surfaces that interact with "foreign" antigenic determinants, constant regions perform the function of
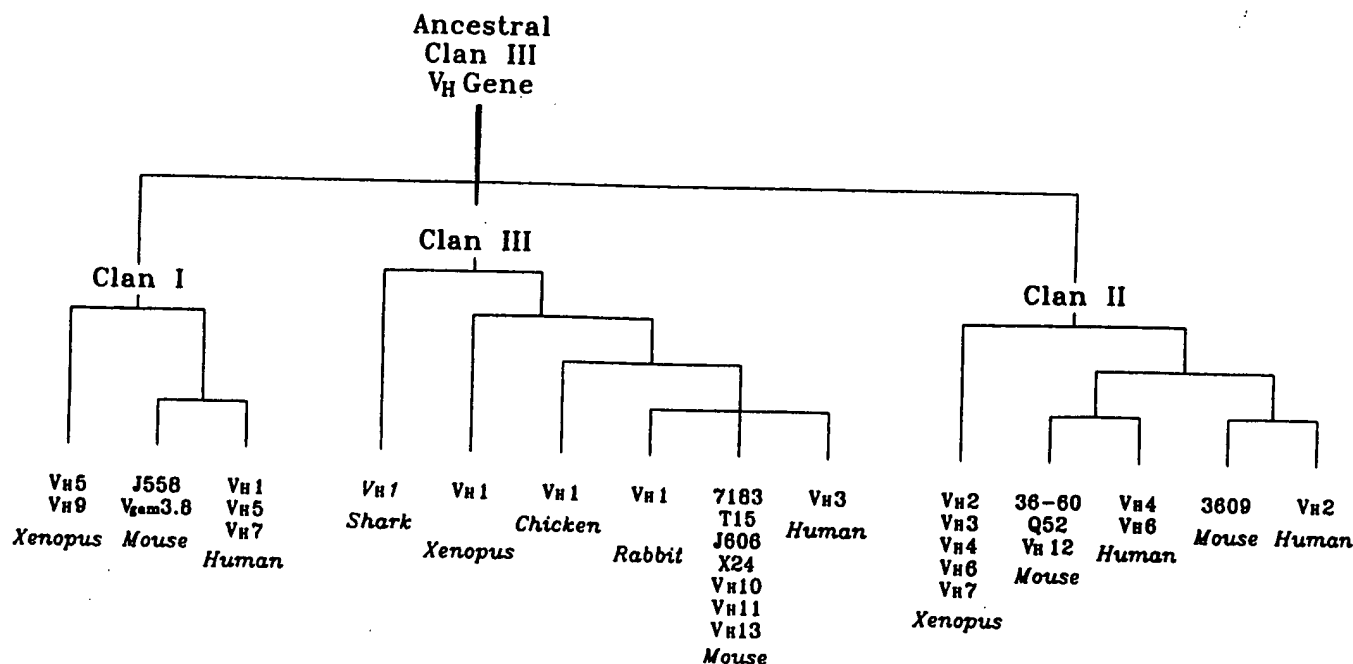
FIG. 7. Clan groupings segregate V$_H$ genes across vertebrate species into distinct clusters of sequences. Note that several different V$_H$ families in a given species can be present within a single clan, reflecting their underlying structural similarity. The lines depicting evolutionary relatedness are not drawn to scale. (From ref. 42, with permission).

```
V_κ1  CAR    DIQMTQSPSTLSASVGDRVAITCRASQNI  SSW    LAWYQQKPGKAPKVLIYKSSSLESGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQYNTFF
      DEE    ---------S----------T--C--G-SV     NKY-N-------------FAA---K--------------------------G-L-----------SY--P
V_κ2  MIL    --VL----LS-PVTP-EPAS-S--S---L  LZ-BGB Y-D--LZ---ZS-ZL---LG-NRA----N-----------B--K--RV-AZ-VGV---N-ALQ-P
      TEW    --V-----LS-PVTP-EPAS-S--S--SL  LH-DGFDY-N--L----QS-ZL---AL-NRA----D-----------K--RVEA--VGV---NZALQAP
V_κ3  CLL    E-V-----A---V-P-E-ATLS-----SV  -NN    --------QP-RL---GA-TRAT-I-A--------E-----R--S----V------NWP
      BL41   E-VL----G---L-P-ESATLS-----SV  --N    --------R-QS-RL--RDA--RAN-I-D-------------I--R-E-----V------S-SP
V_κ4  B17    --V-----DS-AV-L-E-AT-N-KS--SILY--DNKNY--------QP--L---CA-TR-----D-----------------A--V-V-----YNLP
      JI     --V-----DS-AV-L-E-AT-N-KS--SVLY--NNKNY--------QP--L---WA-TR-----D-----------------A--V-V-----D-IP
```

FR1	CDR1	FR2	CDR2	FR3	CDR3

```
V_H1  20P3   QVQLVQSGAEVKKPGASVKVSCKASGYTF  TGYYMHWVRQAPGQGLEWMGWIN PWSGGTNYAQKFGGRVTMTRDTSISTAYMELSRLRSDDTAVYYCAR
      51P1   ---------------------------G--  SS-AIS--------------G-I  -IF-TA------Q-.--I-A-E-T--------S--E---------
V_H2  CE1    --N-RE--PALV-ATHTLTLT-TF--LSVMTR-MSVS-I--P--KA---LAR-D  -DDDKY-GTSLET-L-ISK---KNQVVLIVTNMDPA---T-----
      COR    --T-RE--PALV--TQTLTLT-TF--FSLSS--MCVG-I--P--K----LAR-D  -DDDKY-MTSLET-L-ISK---RNQVVLTMDPVDTA  T-----
V_H3  30P1   E---LE--GGLVQ--G-LRL--A---F--  SS-A-S--------K----VSA-S GSG-S-Y--DSVK--F-IS--N-KN-L-LQMNS--AE---------
      56p1   ----E--GG-VQ--R-LRL--A---F--   SS-A--------K----VAV-S YDGSNKY--DSVK--F-IS--N-KN-L-LQMNS--AE---------
V_H4  71-2   ----QE--PGLV--SETLSLT-TV--GSVSSGS--WS-I--P--K----I-Y-Y  Y--S---NPSLKS---ISV---KN-FSLK--SVTAA---------
      V2-1   ----Q-W--GLL--SETLSLT-AVY-GSV  S---WS-I--P--K----I-Y-Y  Y--S--NNPSLKS-A-ISV---KNQFSLN--SVTAA-----C---
V_H5  251    E-------------E-L-I---G---S-   -S-WTG----M--K------I-Y  -GDSD-R-SPS-Q-Q--NSA-K------LQW-S-KAS---N-----
      32     E-----------E-LRI---G---S-    -S-WIS----M--K------R-D  -SDSY---SPS-Q-N--NSA-K------LQW-S-KAS---N-----
V_H6  15P1   ----Q---PGLV--SQTLSLT-AI--DSVSSNSAAWN-I--S-SR----L-RTYYRSKWYND--VSVKS-I-INP---KNQFSLQ-NSVTPE---------
```

FR1	CDR1	FR2	CDR2	FR3

FIG. 8. Representative sequences from several human V_κ and V_H families. Gaps introduced to optimally align the sequences are indicated by *blank spaces*. Identity between residues in the top sequence and those below it is signified by *dashes*. Canonical boundaries for FRs and CDRs are schematized beneath both the light chain and heavy chain sequence groupings. Note the presence of many "shared substitutions" (amino acids common to two sequences belonging to the same family, but absent in sequences from other families).
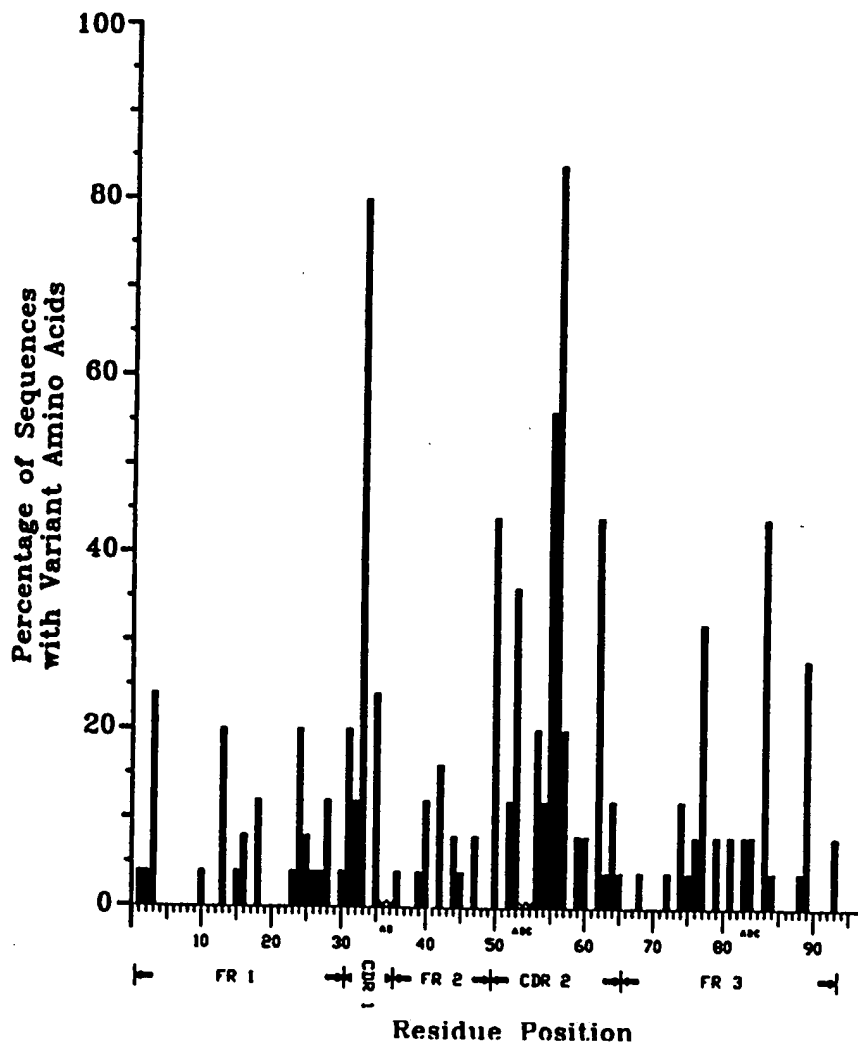
FIG. 9. Sequence variation within a V region family. Twenty-five unique germline sequences from the murine V_H 7183 family were used to generate a consensus sequence for the family. Divergence from the consensus 7183 sequence is represented as the percentage of sequences having a variant amino acid (an amino acid different from the consensus residue) at a particular position. Like the variability plot of heavy chains from multiple families (see Fig. 4), highly divergent positions within the same family also tend to localize to the CDRs. (From ref. 42, with permission.)

binding to the "self" molecules that mediate physiologic and immunologic effector pathways. Light chains have one constant region domain (either C_κ or C_λ) C-terminal to the V_L. The C_L pairs with the first constant region domain of the heavy chain (C_H1) using hydrophobic and disulfide interactions to stabilize heavy and light chain coupling. C_L is not known to specifically bind any other biological moities; therefore, no known effector properties are attributable to light chain C regions. By contrast, all known quaternary associations, biologic characteristics, and physiologic functions of immunoglobulins are governed by heavy chain C regions. Depending upon antibody isotype, heavy chain C regions consist of either three or four C_H domains. The first domain (C_H1) mediates association with C_L as detailed above and is part of the Fab fragment, while the following two (or three) C-terminal domains participate in interchain binding between heavy chain molecules and collectively comprise the Fc. Connection between C_H1 domains of μ and ε antibodies and their respective Fc is accomplished by specialized C_H domains ($C_\mu2$ or $C_\varepsilon2$) that allow for flexibility of the Fabs, yet maintain features of other Ig constant domains. Antibodies of the γ, α, and δ classes, however, utilize a shorter flexible hinge segment for this purpose, which will be specifically detailed elsewhere.

While initial studies indicated that C region sequences were reasonably highly conserved—at least relative to V regions (refer to Fig. 3)—their designation as "constant" is actually somewhat of a misnomer. For instance, studies comparing V and C regions from primitive species have suggested that the lengths of the loops connecting the β strands of constant regions actually vary more than do those of variable regions (51). Furthermore, between different isotypes of the same species, C_H regions share only about 30% sequence identity overall (Fig. 10). C region domains differ in

regard to their interclass homologies such that different domains show various levels of sequence conservation. C_H1 domains are the most similar between isotypes; this may derive from the fact that all share the common function of pairing with immunoglobulin light chains. Similarly, the carboxy-terminal domains of μ, α, and γ ($C_\mu4$, $C_\alpha3$, and $C_\gamma3$) are substantially more related than the average for their Fc as a whole. As x-ray studies have indicated close contact exists between the C-terminal domains of the Fc, this sequence conservation may result from similar constraints as for the C_H1/C_L situation. Moreover, in the case of the μ and α chains, the relatively higher homology likely reflects the common role of the last domain in multimer formation. Also note that the $C_\mu4$ and $C_\alpha3$ domains each possess an 18-amino acid "tailpiece" at their C-termini. The penultimate cysteine residue in each sequence contributes to an intersubunit disulfide bond, allowing IgM and IgA polymerization (52). Much like the V region paradigm, a hierarchy of shared substitutions can be used to distinguish sequences of different domains, classes, and subclasses. In all comparisons between isotypes, however, it is apparent that the majority of conserved residues are localized to the β strands. As was the situation with the V region FRs, these strands are responsible for the proper folding of the domain. In addition, the two cysteines that form the intrachain disulfide bond, and the tryptophan which protects it from solvent reduction, are preserved among all C_H sequences as well.

In analogous fashion to V region evolution, the five heavy chain isotypes likely arose by duplication and diversification of a common gene precursor. This probably occurred in an ancestral organism that preceded mammalian speciation, because examples of all five classes appear in all mammals. Actually, interspecies sequence comparisons of a given isotype demonstrate their similarity far exceeds that of the different isotypes within a species (compare
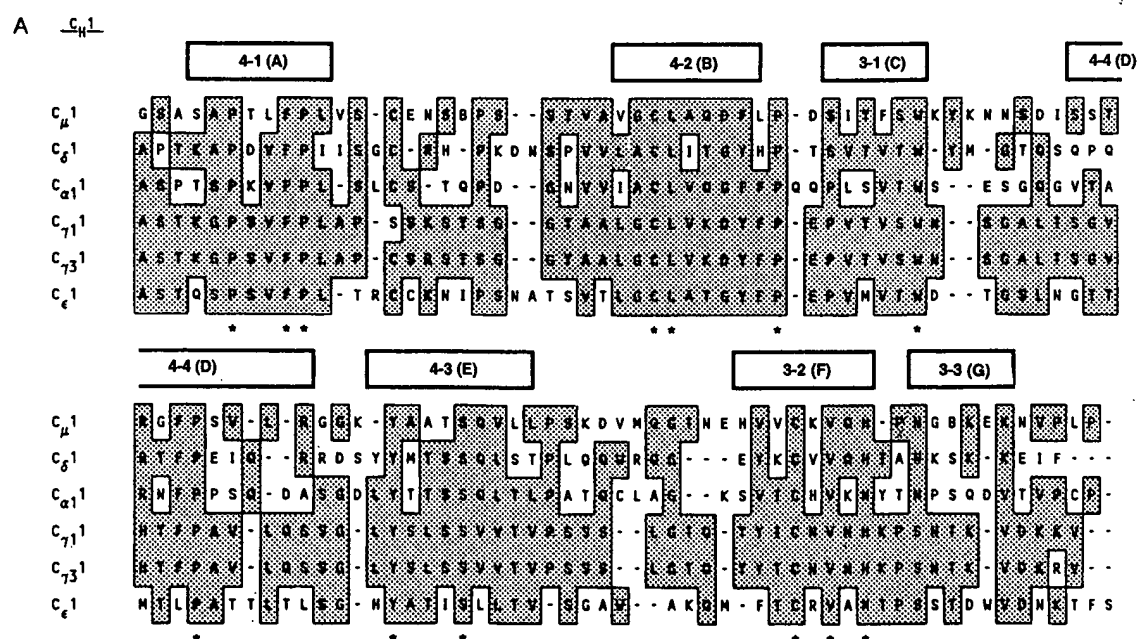
FIG. 10. Comparison of the amino acid sequences of the (**A**) C_H1 and (**B**) Fc regions of all human isotypes (excluding the $\gamma2$, $\gamma4$, and $\alpha2$ subclasses). *Boxed* and *shaded* amino acids represent residues shared by two or more isotypes. *Asterisks* mark amino acids conserved among all six sequences. *Dashes* indicate gaps introduced to maximize homology between sequences. β strands are numbered according to Edmundson and lettered (*in parentheses*) according to Hood in *white boxes* above the alignments.

Figs. 10A and 11). Subclasses represent more recent gene duplication events. Accordingly, distinct subclass profiles exist in various mammalian species. For instance, while humans have two α loci and hence two IgA subclasses (53), murine species have only one α gene, and rabbits have 13 (54)! Due to their later evolutionary divergence, heavy chain subclasses display greater sequence similarity—on the order of 60% to 90%—as evidenced by the γ1 and γ3 alignments depicted in Fig. 10. Despite the high levels of concordance between C regions of related subclasses, even slight differences can have profound functional repercussions. As an example, among the four human IgG subclasses (whose sequences are over 95% identical to one another), IgG1 and IgG3 bind to macrophages and other phagocytes with ease, while IgG2 and IgG4 bind very poorly. This binding is mediated by an Fcγ receptor that has been extensively characterized (detailed later within this chapter). Similar subtle sequence differences are also involved in functional properties affecting immunoglobulin catabolism, placental transfer, and reactivity with staphylococcal protein A (SPA). It is this selective pressure for the ability to perform in a variety of functional capacities that has both maintained the five major C_H classes throughout mammalian species and also driven the evolution of subclasses to their points of divergence in these same
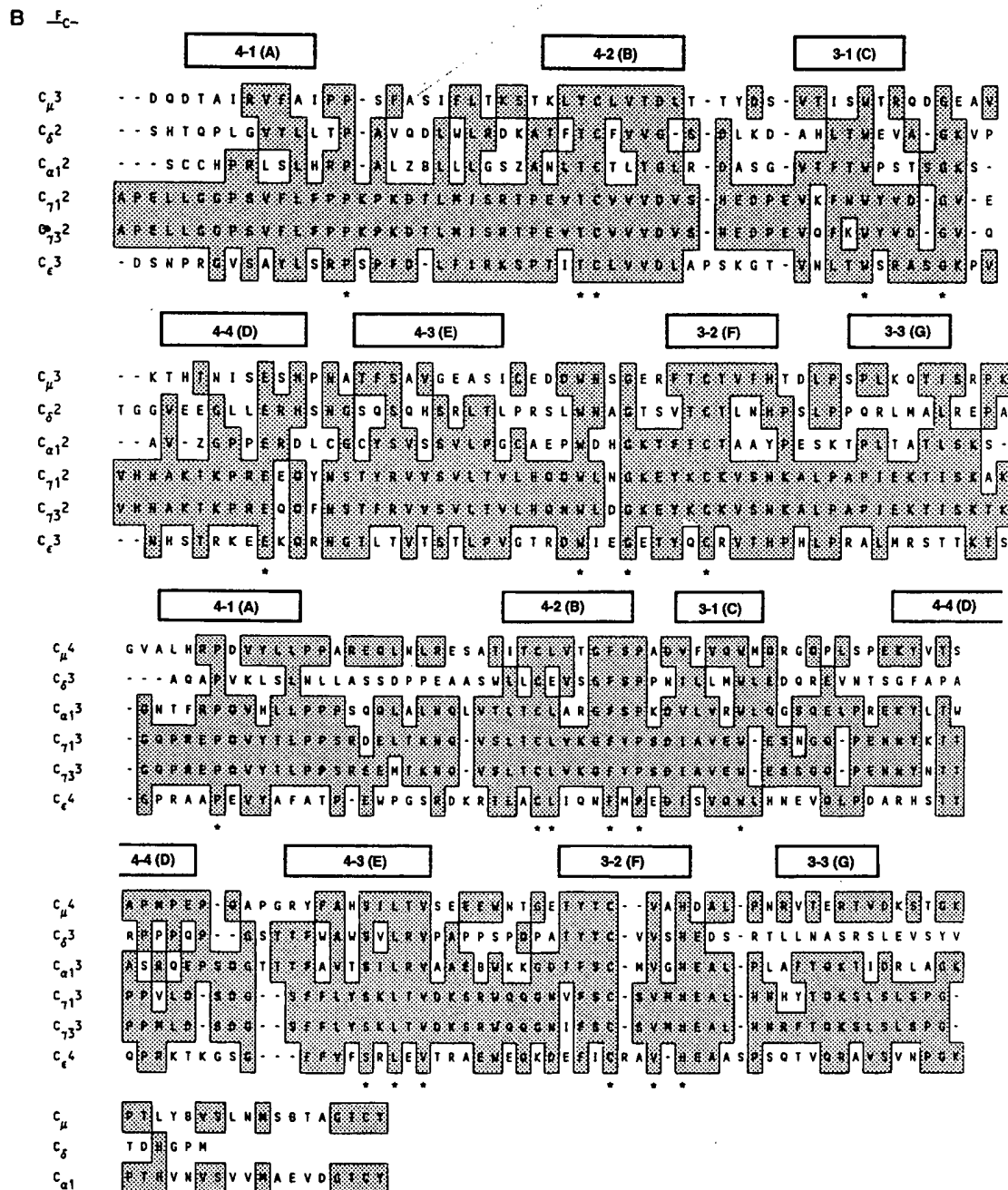
FIG. 10. Continued
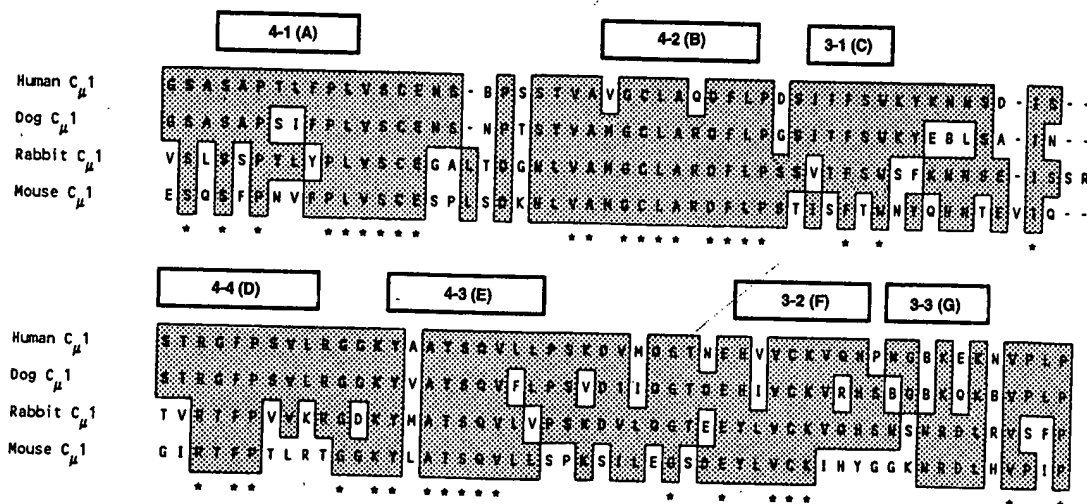
FIG. 11. Protein sequence alignments of the $C_\mu 1$ domains from four mammalian species. Labels are as for Fig. 10. Note the paucity of gapping required and also the high number of absolutely conserved residues (*) across species relative to the alignments of the different human $C_H 1$ isotypes in Fig. 10A.

species. Collectively, these evolutionary changes have bestowed upon the different isotypes the ability to respond to antigenic challenge in a variety of immunologically productive ways.

Despite the absence of evidence to suggest that light chain C regions participate in biologically significant interactions (other than coupling with heavy chains), multiple forms of C_L have also been identified. First, the light chain isotypes C_κ and C_λ—while functionally indistinguishable—exist as separate genetic loci with their own complement of possible V and J gene segment partners. This is in contrast to the situation with the heavy chain locus, in which any rearranged VDJ can potentially become associated with any C_H isotype via the process of class switching (see Chapter 24). While the use of κ versus λ isotypes seems to be inconsequential as to the antibody's efficacy, differences in their utilization are nonetheless present. In human immunoglobulins the κ:λ ratio is approximately 70:30, while in murine systems about 95% of antibody is of the κ class (20). The reasons for these imbalances are yet to be definitively explained, but are probably related to the number of V_κ and V_λ genes available for use in the respective genomes of these organisms. As was the case for heavy chain C regions, light chain isotypes are well conserved across species (Fig. 12). In addition, within a species, κ and λ classes are more similar to each other (about 38%) than were the different C_H isotypes (compare Figs. 10 and 12). Note also the presence of a terminal (C_κ) or penultimate (C_λ) cysteine residue in these sequences. This half-cystine is responsible for the light chain's contribution to the H–L interchain disulfide bond. A second point of variation within C_L regions concerns C_λ subclasses. While only a single C_κ gene is present in human and mouse, five or four λ-chain subclasses exist in human or murine genomes, respectively. A final deviation from "constancy" involves the C_κ domain. While only one κ constant region gene is found in the locus, three human κ allotypes have been identified. These alleles differ in regard to their residues at the surface-exposed positions 153 and 191, such that these allotypic markers are able to serve as antigenic determinants (55).

As mentioned previously, hinges—either as distinct domains in the cases of the μ and ϵ chains or as shorter specialized segments for α, γ, and δ antibodies—connect the Fab and Fc portions of the

immunoglobulin molecule. However, the lack of extensive hinge structure in nonmammalian heavy chain sequences indicates that the evolution of this function largely occurred after mammalian radiation (51). Consequently, discussion here is restricted to mammalian hinges as typified by human sequences.

Hinge regions not only connect Fab to Fc, but also contain the disulfide bonds that covalently link the two heavy chains (discussed further in the section on quaternary structure) in the Fc portion of the antibody molecule. They display great variability between isotypes and are generally encoded by unique exons. For example, the hinges of human IgG1, IgG2, and IgG4 are each produced from a single short exon encoding a peptide of between 12 and 15 amino acids. Alternatively, the IgG3 hinge derives from four distinct exons, resulting in a hinge region that spans approximately four times as many residues as the other γ isotypes (56). Owing to similarities between the different hinge sequences and the extra C_μ and C_ϵ domain sequences, it is thought that hinges evolved from the $C_{\mu/\epsilon} 2$ domain; unfortunately the hinge sequences are too short and the homologies too weak to trace hinge lineage with certainty. It is clear from sequence comparisons, however, that the $C_\gamma 3$ hinge evolved from the $C_\gamma 1$ hinge by multiple duplication (further substantiated by the aforementioned $C_\gamma 1$ and $C_\gamma 3$ hinge exon arrangements). Figure 13 aligns the extra domains of C_μ and C_ϵ, in addition to the hinges of several other human isotypes.

In addition to the variable, constant, and hinge regions of antibodies, other primary sequence features must also be considered. For instance, both heavy and light chains of immunoglobulin are synthesized with a leader peptide (almost entirely encoded by a separate exon upstream of the V region) of 16 to 26 amino acids in length. During translation, this leader (more generally referred to as a signal peptide) directs the mRNA/ribosome/polypeptide complex to the rough endoplasmic reticulum (RER) where translation is completed. During the synthesis and extrusion of the nascent polypeptide through the RER membrane, the signal peptide is then removed by specific proteolytic cleavage.

Finally, for a comprehensive discussion of primary immunoglobulin structure, mention must be made of sequences present exclusively in the case of surface immunoglobulin expression. Heavy
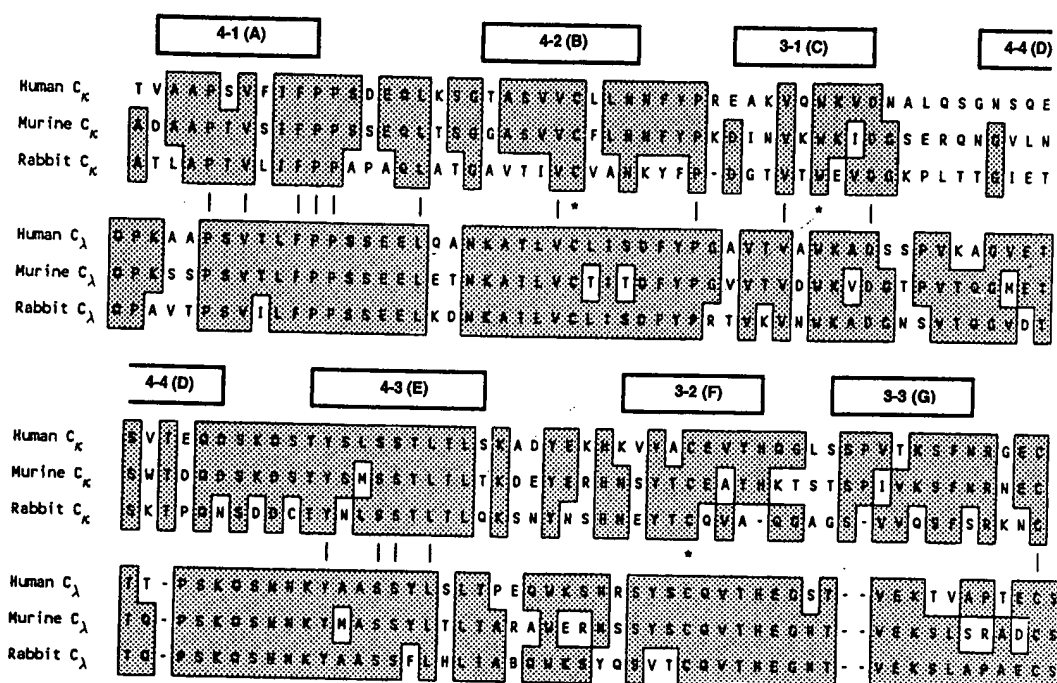
FIG. 12. Comparison of the C_κ and C_λ domains from three mammalian species. Labeling is as for Figs. 10 and 11, except *vertical lines* now represent amino acids shared among all six sequences and *asterisks* specifically denote the invariant tryptophan and cysteines at the core of the domain (also shared by all sequences). As in Fig. 11, little gapping is needed and high homology persists across species.

A) Epsilon and Mu C_H2
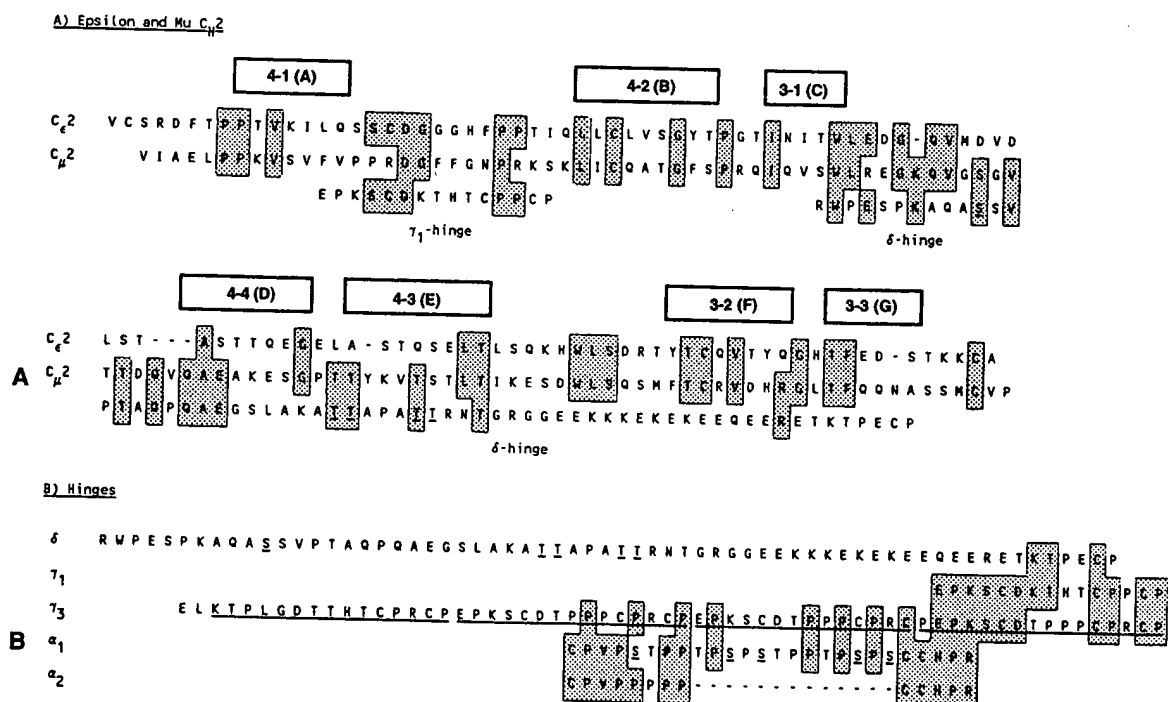
δ-hinge

B) Hinges

FIG. 13. Alignment of the amino acid sequences of (A) the human $C_\mu2$ and $C_\epsilon2$ domains and the γ1 and δ hinges and (B) several human immunoglobulin hinge regions. Labeling is as for Figs. 10–12. In (A), the $C_{\mu/\epsilon}2$ domains are compared with the hinges of two different isotypes to display their potential, although limited, evolutionary relationship. Part (B) demonstrates several features of hinge regions. Note the high proline, cysteine, and serine/threonine content of the hinges, which consists of the bulk of the homologies between isotypes. This interclass homology is reasonably low, whereas intersubclass homologies (compare γ1 to γ3 and α1 to α2) are considerably higher. Note the fourfold duplication of the γ1 hinge in γ3 (the repeating unit is *underlined*). Also note the deletion within the α2 hinge relative to α1. O-linked glycosylation sites in the δ and α1 hinges are also *underlined*.

chains of all isotypes can exist either as secreted antibody or as membrane-bound immunoglobulin (mIg), which serves as the central component of the antigen-specific B cell receptor (BCR). The choice between mIg and secreted immunoglobulin is manifest at the level of alternative mRNA splicing at the 3' end of the message. Differential processing in favor of the surface immunoglobulin form results in the replacement of the hydrophilic carboxy-terminal residues of secreted antibody with a stretch of hydrophobic amino acids (which anchors the immunoglobulin in the cell membrane) and a short cytoplasmic tail (57,58). Expression of these sequences not only tethers the immunoglobulin to the cell surface, but also governs the ability of mIg to interact with other constituents of the BCR necessary for signal propagation and eventual activation by antigen (see Chapter 7). As such, the regulation of this splicing event and its ultimate protein products is tightly controlled through the stages of B lymphocyte differentiation so as to ensure the proper production of membrane immunoglobulin versus secreted antibody at the appropriate developmental stage of the B cell (see Chapters 5 and 6).

Secondary Structure—The Immunoglobulin Fold

With the exceptions of the hinge and cytoplasmic tail, the properties of the immunoglobulin fold as a protein motif dominate all aspects of immunoglobulin structure, from primary to quaternary. In regard to secondary structure, this refers to the different patterns of β-pleated sheet formation assumed by the V and C regions. As explained earlier, all immunoglobulin folds are composed of two layers of antiparallel β sheet arranged in a sandwich (or β-barrel) that encloses a hydrophobic interior. Early on, it was recognized that each immunoglobulin domain contains seven polypeptide β strands, four of which comprise one β-pleated sheet, the other sheet consisting of the remaining three strands. Accordingly, a nomenclature was developed that reflected within which β sheet (four-stranded or three-stranded) a particular strand was located, using numbering (3-1, 3-2, 3-3 and 4-1, 4-2, 4-3, 4-4). Subsequent stud-

ies revealed that superimposed upon this secondary structural organization shared by all immunoglobulin folds was a discrepancy between V and C region immunoglobulin folds; some of the "extra" amino acid residues found in variable regions actually participated in β sheet formation, giving rise to two additional β strands. Consequently, the immunoglobulin folds of V domains actually form a barrel using five-stranded (analogous to the C region three-strand sheet) and four-stranded β-pleated sheet layers. A second naming system using letter designations (A, B, C, C', C", D, E, F, and G) for the different strands of immunoglobulin fold β-pleated sheets makes allowance for these extra β strands. Figure 14 displays "unfolded" immunoglobulin folds of each type so as to schematically represent the secondary structure of both V and C immunoglobulin domains. Note that while the letter nomenclature proceeds in accordance with the primary structure of the protein (A, B, C, D, etc.), the numbering system of β strands is nonlinear with respect to the primary sequence (4-1, 4-2, 3-1, 4-4, 4-3, 3-2, 3-3). Rather, this naming system is designed so as to coincide with the three-dimensional orientation of the strands in the context of the fully folded immunoglobulin domain (refer back to Fig. 2), which will be discussed in greater detail below.

Tertiary Structure—The Immunoglobulin Domain

Logically, if the β sheets of the immunoglobulin fold are the dominant secondary structural protein motif of the antibody, then the fully-folded tertiary structural correlate of this paradigm is that of the immunoglobulin domain. As has been explained, Ig domains are founded upon the premise of two layers of β-pleated sheet, sandwiched around a core of hydrophobic side chains, to form a compact globular structure. Of course, upon this general structural framework, Ig domains vary considerably. Still, certain features common to all Ig domains of actual immunoglobulin molecules (as opposed to Ig domains of other IgSF members) are invariant. Central to the Ig domain, both literally and figuratively, is the presence of the two cysteines that form the intradomain disulfide bond and
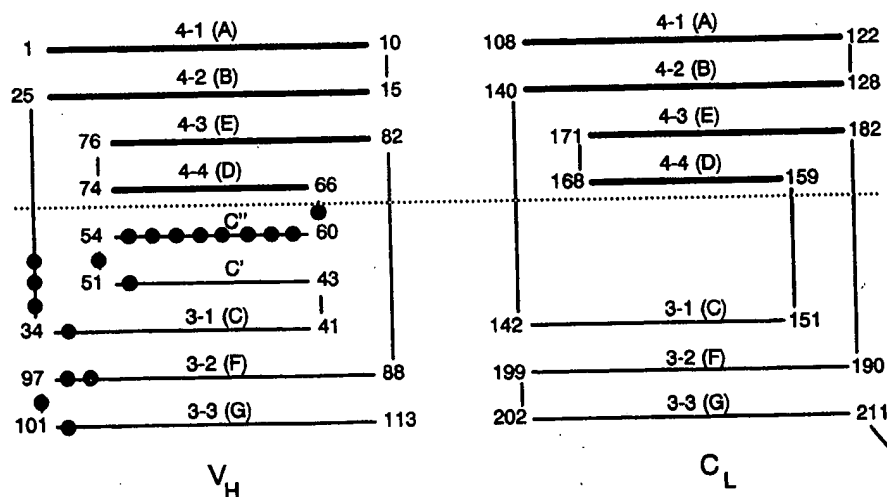
FIG. 14. Schematic diagram of the secondary structure of "unfolded" immunoglobulin domains. β strands are represented by *horizontal lines,* loops connecting β strands are depicted as *vertical lines. Bold horizontal lines* are in the four-stranded face of the domain, *light horizontal lines* are in the three- (five-) stranded β sheet. *Dotted regions* denote CDRs. Strands are numbered according to Edmundson and lettered (*in parenthesis*) according to Hood. In three dimensions, immunoglobulin domains are folded with the *light horizontal strands* under the *bold horizontal strands* using the dotted line as an axis.

the tryptophan which protects this bond from hydrolysis. All domains of immunoglobulin—variable region or constant, heavy chain or light—conserve these three key residues, and they occupy homologous positions in the different domains of the fully-folded protein.

Beyond these key core residues, however, different Ig domains are still able to maintain similar tertiary structures in the face of dramatic differences in their primary sequences. This chiefly derives from the fact that the identity of a particular residue at a particular position is not nearly so important to proper Ig domain folding as is the character of the particular residue at that position. In other words, as long as amino acids with side chains compatible with β-pleated sheet formation are present in the proper locations and those necessary to terminate β strands are similarly in the correct places, their actual identity appears not to be crucial.

There are, of course, other residues essential to proper folding and functioning of Ig domains, but these are specific to particular domains and not common to all Ig domains within immunoglobulin. For example, the FR4 motif W/F-G-X-G is widely conserved, but only among V regions, where it serves to create a "β bulge" necessary for proper V_H/V_L dimerization. Similarly, the V_H-specific (G-L-E-W-hydrophobic) and V_L-specific (P-hydrophilic-hydrophobic-L-hydrophobic) FR2 sequence motifs—and their accompanying β bulges—are also conserved tertiary structures among their respective subsets of Ig domains. Finally, as was discussed in secondary structure, V and C domains also differ in terms of their basic arrangement of the immunoglobulin fold such that V domains are composed of a four-strand (A, B, E, and D) and a five-strand (C", C', C, F, and G) layer, whereas C domains consist of four-stranded (A, B, E, and D) and three-stranded (C, F, and G) β sheets. Clearly, this distinguishes variable and constant domains at the tertiary structural level as well.

Once again, as with primary and secondary structural considerations, examination of tertiary structure can be facilitated by distinguishing between V and C domains. In the case of variable regions, whether from heavy or light chains, obviously no single protein structure will ever fully suffice in describing the entire group. This is because each domain, V_H or V_L, is in effect a new structure, and unless solved crystallographically, can only be postulated. Nonetheless, two broad generalities as pertains to V region tertiary structure are evident. First, the similarity between two different V domain structures tends to closely parallel their relatedness at the genetic level; that is, one can reasonably predict that two different V_H structures will be more similar to each other than to a V_L structure, two V_H domains belonging to the same clan will be more similar than to one from a different clan, and so on. Colorplates 1 and 2, which compare molecular models of FR1 regions from antibodies belonging to the same and different clans, are particularly compelling in this regard. While undoubtedly exceptions exist where two proteins are genetically similar but diverge at a few key residues with important structural consequences, in most instances this rule is valid for extrapolating the tertiary structures of unsolved V domains. In this context, it is perhaps also important to note that the FR1 region utilized to assign clan identity is solvent exposed and distal to the antigen-binding site, while the FR3 region which correlates best with family designation is immediately adjacent to the binding site, and capable of affecting its conformation and even interacting with antigen directly (42). In light of several reports linking over-representation of certain families in the repertoires reactive against particular antigenic specificities, the tertiary

structural predictions available by this means perhaps take on added significance.

The second pervasive trend which is apparent upon scrutiny of V region tertiary structures is the tendency of the antigen-binding site to represent a "nested gradient of antibody diversity" (42). Recall, investigations into the very first antibody proteins identified nonconserved "hypervariable" regions and reasonably well-conserved "framework" regions, which were in turn postulated to correspond to the antigen-binding and structural foundations, respectively, of the molecule. These hypotheses proved correct, as the FRs were demonstrated to coincide with the β strands, and the CDRs were shown to derive from the variable loops that interconnect the strands. Moreover, current understanding of antibody structure makes it possible to recognize that, as a general rule, the most variable residues of an immunoglobulin V region localize immediately proximal to the antigen-binding site, whereas those that are most conserved tend to be distant to that site. Colorplate 3 provides one such example of this concept. Thus, the three-dimensional context in which amino acids interact to create a platform for ligand-binding (paratope) diverges dramatically from antibody to antibody.

There are several factors that influence the tertiary structure of the paratope itself, and their composite effect can be complex. First, sequence variation of two types in the loops can profoundly alter antigenic specificity and affinity. CDRs vary considerably in length, both as a function of V gene usage (affecting primarily CDRs 1 and 2) and as a consequence of junctional diversity (affecting only CDR3). Second, CDRs obviously differ significantly in terms of their sequence composition, due once again to gene usage and junctional diversity, and in addition to somatic hypermutation. In this way, diversity-generating mechanisms target amino acid variability to the CDR loops where they are most apt to change both the physical shape and chemical nature of the combining site. Also, because FR residues near CDR boundaries frequently interact with antigen directly (59), alterations in these positions affect structural variety as well. Even glycosylation of CDR asparagine residues has been implicated in changing loop conformation and antigen binding (60,61). Conformational variability can also play a critical role in diversifying the paratope surface, as CDR loops have been shown to interact extensively with adjacent FR amino acids and with each other (62,63). These studies, and failed attempts to engineer antibodies by simply swapping CDRs onto different FR backbones without appreciably affecting affinity for antigen, have further demonstrated that while V regions can be conveniently dissected into FRs and CDRs at the primary structural level, in actuality these elements cooperate to facilitate antigen-binding rather than acting as discrete elements.

While the limitless capacity of the immune system to generate new variable region domains makes impossible the absolute elucidation of all potential V region structures, the relatively smaller number of constant regions has allowed reasonable progress to be made in terms of assigning definitive tertiary structures to the C region domains. To date, x-ray diffraction analysis has resulted in a high resolution structure for only the Fc fragment of IgG1 (11). An α-carbon backbone of this structure is shown in Fig. 15A. Additionally, one whole immunoglobulin (an unusual IgG1 molecule with a hinge deletion) has been crystallized (64). Other Fc isotypes, based on the Fcγ1 paradigm, can be modeled (see Fig. 15B) using sequence homology with IgG and energy minimization calculations (65). Otherwise, almost all three-dimensional studies have utilized Fab fragments or Fv fragments, often produced in

A

B
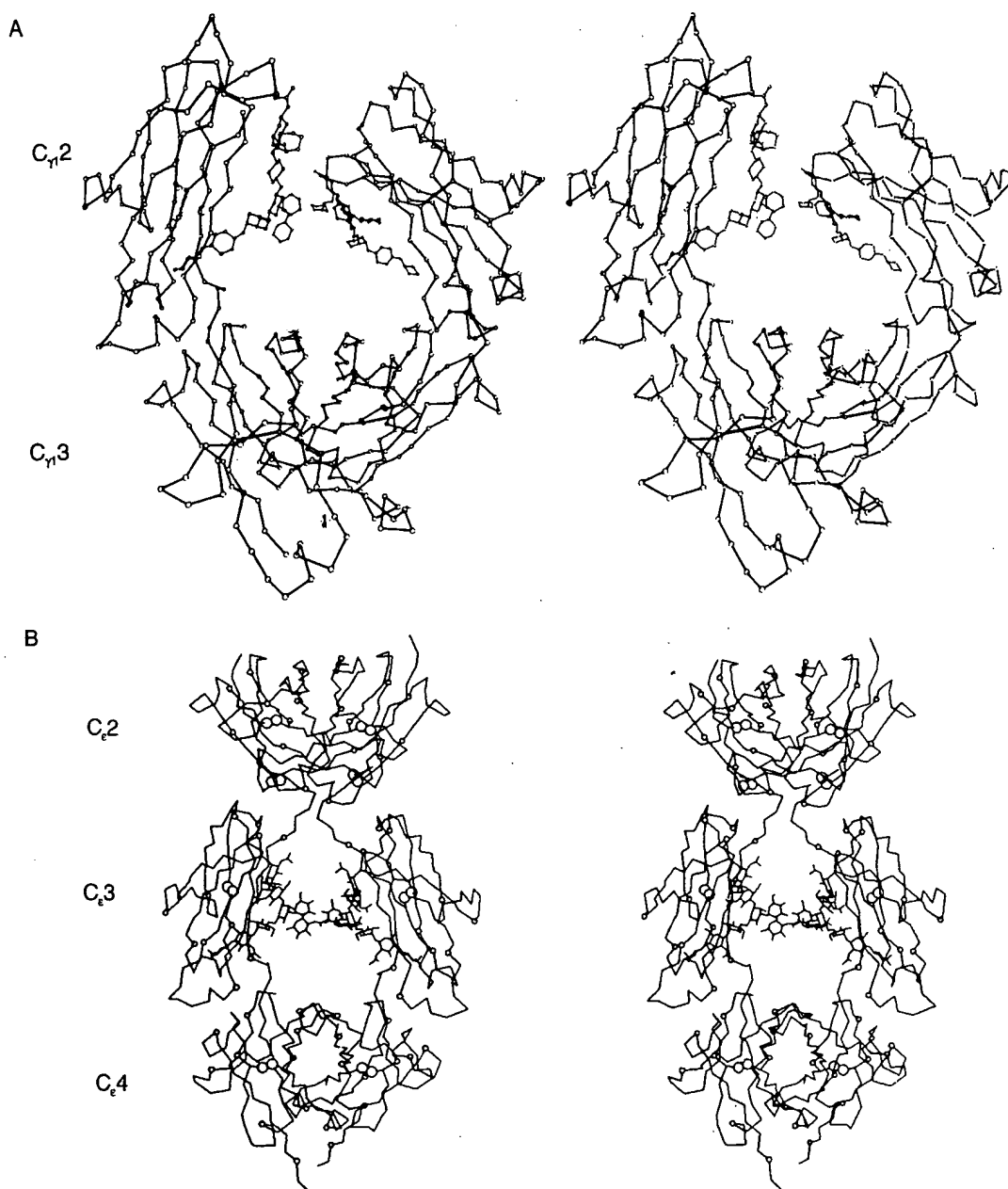
FIG. 15. Stereoviews of the α-carbon backbones of (A) Fcγ1 and (B) Fcε. In both cases, the C-terminal domains dimerize, but the penultimate domains ($C_\gamma 2$ or $C_\varepsilon 3$) contact their counterparts only via their carbohydrate moieties. The Fcε structure is a prediction modeled on the IgG1 crystal. As IgG has no $C_\varepsilon 2$ domain equivalent, the $C_\varepsilon 2$ structure is less reliable conjecture than is the rest of the molecule. (From refs. 12 and 65, with permission.)

bacteria. Of note, the IgG Fc crystal has also been solved as a co-crystal with staphylococcal protein A (SPA) (12). Importantly, this structure reveals the binding of SPA between both the $C_\gamma 2$ and $C_\gamma 3$ domains. This is contrary to the notion originally promulgated that single domains would perform IgG functions. Another example of sharing of function between domains is the Fcα receptor binding site on human IgA, which also involves both $C_\alpha 2$ and $C_\alpha 3$ (66). Surprising results such as these demonstrate the need for continuing investigation—by both crystallographic and other means—into the structural intricacies of all of the immunoglobulin isotypes. In

any case, in the intervening time, the general properties of Fc topology, though proven only for IgG, are reasonably generalized to the other isotypes, since the residues involved in the various contacts are largely conserved.

As for the V region domains, constant region domain structure is governed by general principles that tend to hold true in most examples of C region domains thus far studied. The overriding consideration in this regard, as for the V domains, concerns the relative concentration of variability outside the β strands. C regions conserve a much higher percentage of residues from domain to
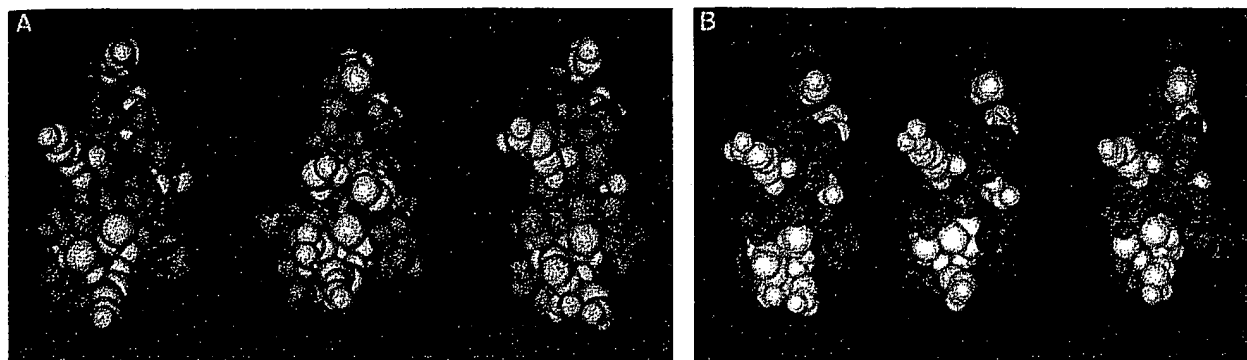
COLORPLATE 1. Three-dimensional models of immunoglobulin FR1 regions. Space-filling representations of amino acids 6–24 are displayed with orientation such that the CDRs would be above, and the C_H1 domain below, the models. Amino acid residues 6, 9, 12, 13, 16, 19, and 23 are colored *yellow* for reference. In (**A**), antibodies deriving from each of the three clans—clan I HyHEL-5 (*left*), clan II NEW (*middle*), and clan III KOL (*right*)—are presented. Clearly these structures differ in their structural characteristics. In (**B**), three clan III antibodies—human V_H3 family KOL (*left*), murine V_HT15 family MCPC603 (*middle*), and murine V_HX24 family J539 (right)—are compared. Note the similarity of these three FR1 loops relative to those in part (**A**). (From ref. 41, with permission.)
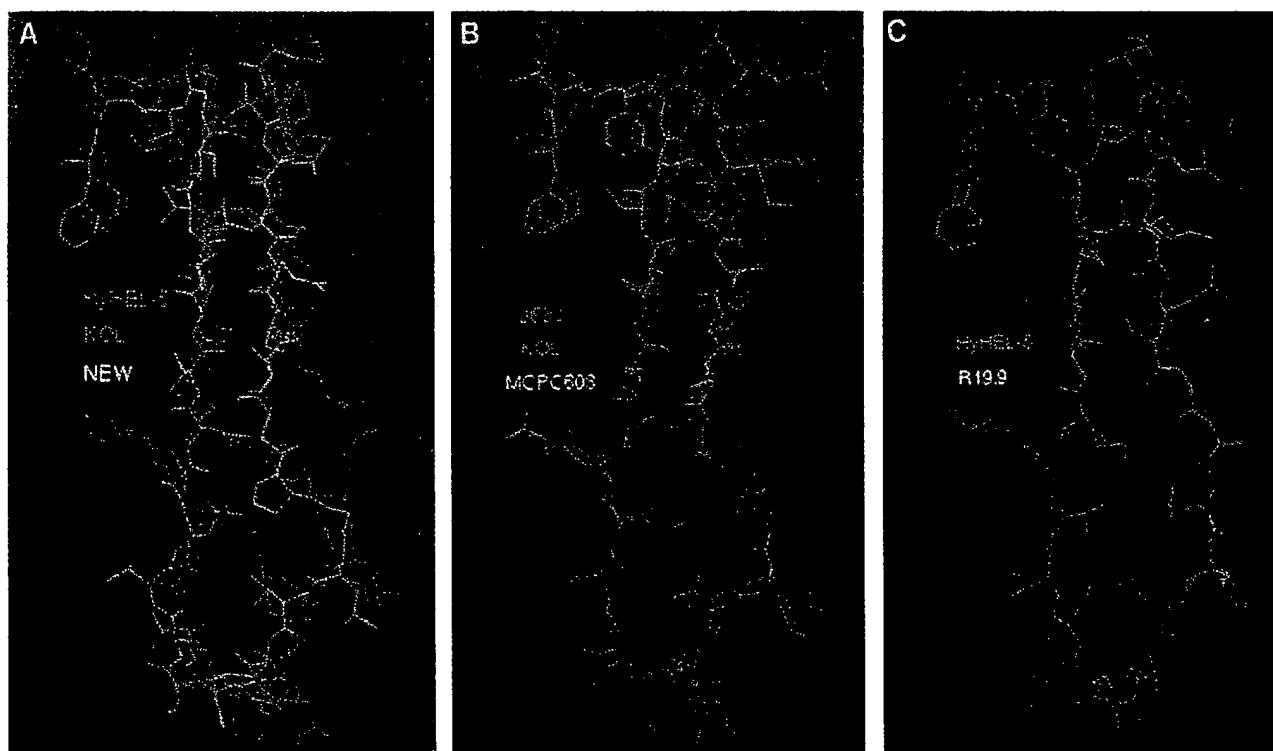
COLORPLATE 2. Superposition of FR1 regions from antibodies of (**A**) different, and (**B,C**) the same, clans. Stick diagrams of amino acid residues and their respective side chains are overlayed to facilitate comparisons. In (**A**), clan I HyHEL-5 and clan II NEW are superimposed on clan III KOL. These molecules can be seen to vary significantly. In (**B**), clan III antibodies MCPC603, J539, and KOL are modeled. Note that the agreement between structures even extends to side-chain sizes and orientations. In (**C**), clan I R19.9 antibody is superimposed on clan I HyHEL-5 with similar results. (From ref. 41, with permission.)
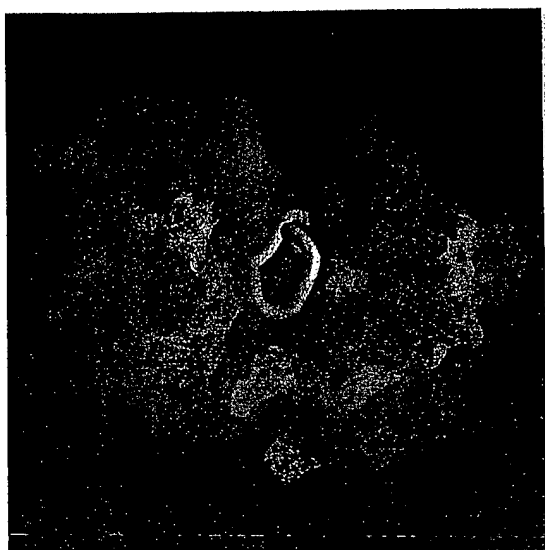
(a)
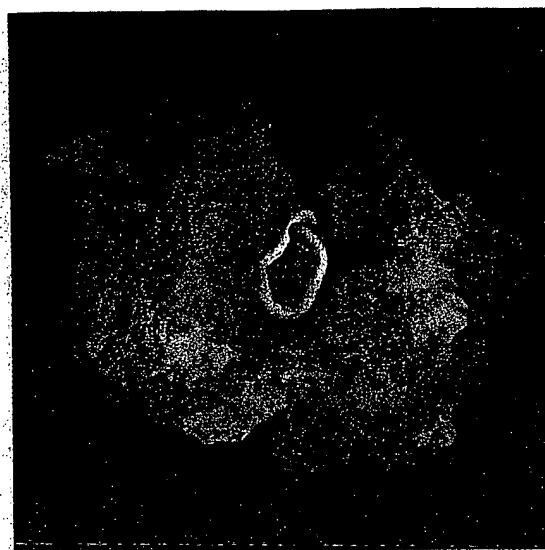
(b)

(c)

(d)

COLORPLATE 3. The antigen-combining site is the product of a nested gradient of diversity. Antibody sequence variation is mapped on the template of the surface of the antibody POT using a scale of *blue* (highly conserved) to *red* (highly divergent). The V_κ domain is to the left of each model, and the V_H domain is on the right. The most highly variable CDR3 loop of the V_H is depicted as a *gray ribbon* in the center of the diagrams; the highly variable V_κ CDR3 is not shown in these representations. In (A), germ-line diversity is displayed. In (B), diversity introduced by somatic hypermutation is presented. In (C), the sum of these two diversities is depicted. Note that in all cases, variation is predominantly restricted to the antigen-binding site, and diversity is highest in close proximity to the V_H CDR3. In (D), residues that have been demonstrated to make direct side-chain contacts with antigen in 21 separate crystals are plotted on the same *blue* (zero contacts) to *red* (up to 16 contacts) scale. Thus, the presence of diversity and the tendency to contact antigen are intimately related. (From ref. 58a, with permission.)
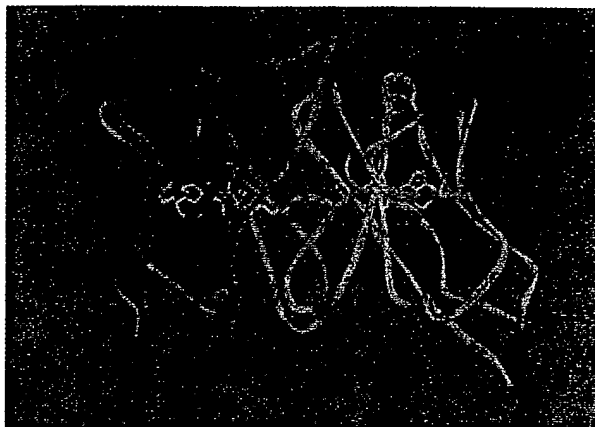
COLORPLATE 4. Ribbon diagram of the α-carbon backbone of an immunoglobulin Fv fragment. The heavy chain is shown in *blue*, and the light chain is represnted in *violet*. The invariant cysteines, tryptophans, and aromatic residues at the core of the V_H and V_L domains are shown in *yellow*. Orientation of the Fv is such that the antigen-binding site is at the *top*, and the C_{HI}/C_L domains would lie at the *bottom*, of the plate. (From ref. 42, with permission.)
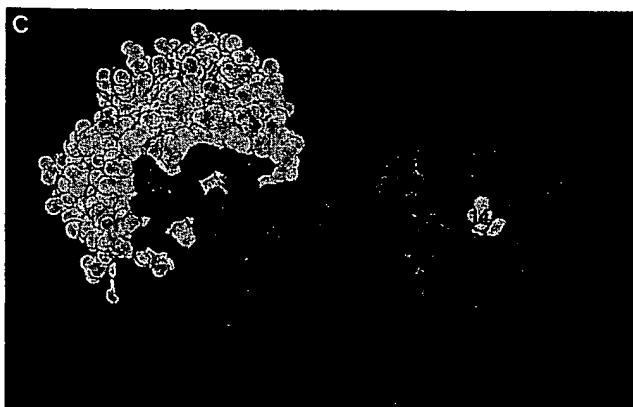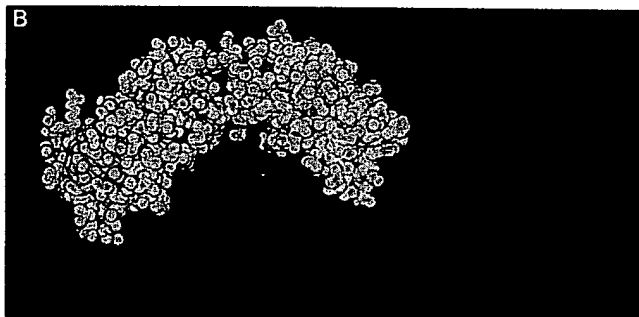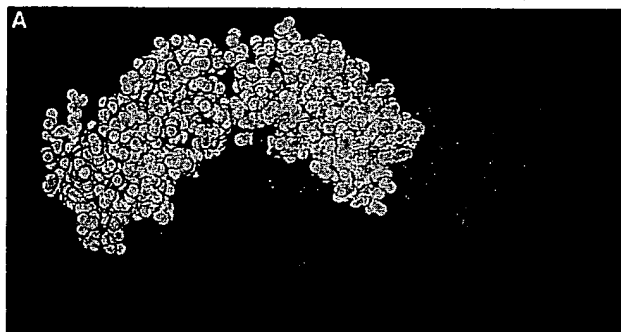
COLORPLATE 5. Model of the lysozyme–antilysozyme antibody (Fab D1.3) complex. The lysozyme molecule is depicted in *green*, and its glutamine-121 residue in *red*. The D1.3 heavy chain is shown in *blue*, and the light chain in *yellow*. In (**A**), the complex is seen as it was in the crystal—as a blunt, end-to-end interface of the two molecules. In (**B**), the two proteins have been separated to demonstrate their structural complementarity. In (**C**), the molecules have been rotated toward the viewer to allow visualization of important contact amino acids (now portrayed in *red*, Gln-121 in *violet*). (From ref. 109a, with permission.)
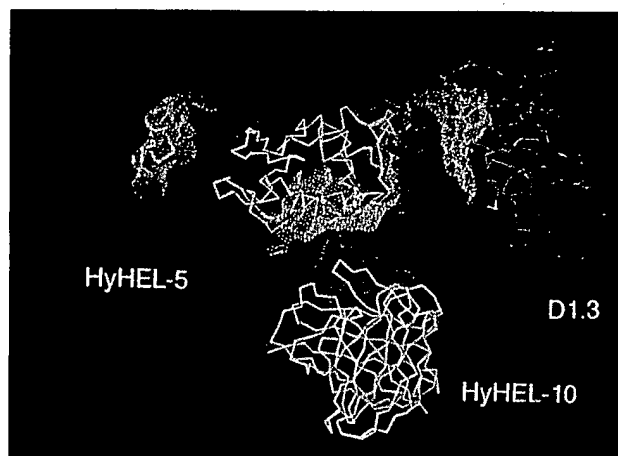
COLORPLATE 6. Model of lysozyme and three antilysozyme complexes. The Fab D1.3 (see Colorplate 5) is included for reference along with two additional antibodies: HyHEL-5 (*left*) and HyHEL-10 (*below*) Molecules are shown as α-carbon backbones except for colored van der Waals surfaces involved in binding. (From ref. 110 with permission.)
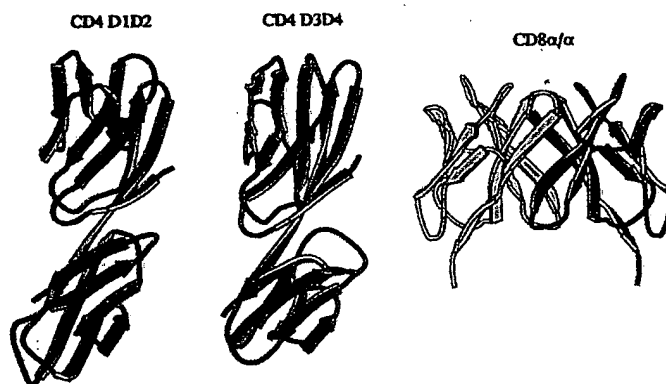
COLORPLATE 7. Ribbon diagrams of CD4 D1D2 (*red*), CD4 D3D4 (*green*), and a CD8α/α homodimer (one subunit *yellow*, the other *blue*). Note the continuous β strand that links domains D1 to D2 and D3 to D4 in CD4. This causes the D1D2 and D3D4 segments to be rigid structurally. (From ref. 207, with permission.)

domain—across immunoglobulin class and even across species—than do individual V regions (refer back to Figs. 10–12), but even so, stretches of relatively lower identity can be localized to distinct parts of the protein. Even though C domains are not formally delineated into framework and hypervariable regions, the tenets used to classify the subdomains of V regions are still applicable when discussing C regions. In Fig. 10A, for example, which compares different human C_H1 domains, note the areas of not only lower conservation but also where gapping is necessary to preserve alignment. Without exception, these regions are most prevalent between strands, especially the loops connecting strands 4-1 and 4-2 (A and B), 4-2 and 3-1 (B and C), 3-1 and 4-4 (C and D), 4-4 and 4-3 (D and E), and 4-3 and 3-2 (E and F). Conversely, the areas of highest conservation are found within the β strands (especially 4-1, 4-2, 3-1, 4-3, and 3-2), where, in addition to the residues needed for the intradomain disulfide linkage present in all Ig domains, the amino acids necessary for main-chain folding, stabilization, and dimerization of the domain reside.

Therefore, as was the case for V regions, the least conserved segments of antibody C domains coincide with the loops that interconnect the different β strands in the fully-folded protein. In the case of C region domains, this refers to divergent loops found at each end of the Ig domain, unlike V regions, where the CDRs are clustered at one end of the domain. Not surprisingly, these loops are also where the preponderance of functional interactions have been localized in binding studies. For instance, the binding site on IgG for the $Fc\gamma RI$ receptor is located near the hinge region in just such a loop of the $C_\gamma 2$ domain. Likewise, the previously mentioned binding of SPA with IgG Fc involves similar loops in both the $C_\gamma 2$ and $C_\gamma 3$ regions (12). In another example, although no crystal structure is available for IgA, by modeling the IgA sequence on the IgG Fc three-dimensional structure, an exposed loop in the $C_\alpha 3$ domain is predicted to be the binding site for the polymeric immunoglobulin receptor.

Thus, the areas of greatest divergence between different C domains are also those implicated in mediating the different biological effects that distinguish one class or subclass of antibody from another. It is fair to speculate that, like the pressures driving the evolution of diversity-generating mechanisms in V regions, similar forces have used the template of the C region Ig domain to select for a variety of distinct binding capabilities and functional attributes, and have utilized a similar region of the domain for these purposes. In a manner analogous to that seen for variable region tertiary structure, those parts of the C region domain that are most malleable structurally are the very same selected evolutionarily for the acquiescence of new biological characteristics.

Quaternary Structure—The Immunoglobulin Monomer

Although, as stated previously, immunoglobulin domains are each capable of folding independently into stable tertiary globular structures, neither individual Ig domains alone nor entire heavy or light chains are ever encountered under normal physiological circumstances. Rather, the simplest form of this protein that occurs naturally is that of the immunoglobulin monomer schematically depicted in Fig. 1. The complete "monomeric" antibody molecule is actually a four-chain dimer of a heterodimer covalently linked by multiple interchain disulfide bonds. In almost all cases, the heavy and light chains are joined by a single cystine to form a "half-

monomer" with one complete antigen-binding site, and one or more disulfide bonds in the hinge regions (or hinge domains) link the two heavy chains to form the bivalent tetramer.

Figure 16 represents in two dimensions a member of each of the five classes of human immunoglobulin. Notice that the primary differences between these structures are the presence of the extra C domain in the IgM and IgE isotypes and the number and placement of disulfide linkages and carbohydrate derivatives among the different molecules. Like the intrachain disulfide bond considered in tertiary structure, the interchain cystine attaching heavy and light chains is highly conserved. The cysteines participating in this bond are located at the N-terminal end of C_H1 and the C-terminal end of C_L. IgG1 is an exception where the cysteine donated by the heavy chain is found at the carboxyl end of C_H1 (55). Another exception to the typical bonding pattern is found in the A2m(1) allotype of IgA2. Distinctively, this particular isotype utilizes H–H and L–L disulfide pairing to stabilize the four chains instead of the usual H–L linkage (67). The interchain cystines joining the heavy chains are more variable in both number and position between different isotypes. Generally, H–H bonds are formed between the hinge regions or, in the cases of IgM and IgE, analogous positions in $C_\mu 2$ or $C_\epsilon 2$ domains, respectively. In addition, both IgM and IgA have two additional half-cystines that have dual roles, depending upon whether the antibody is incorporated into a polymeric form of immunoglobulin. An extra cysteine near the C-terminus is involved in a disulfide bond occurring during J chain–mediated polymerization; another half-cystine in the $C_\mu 3$ or $C_\alpha 2$ domain forms an inter-subunit cystine in the case of IgM or IgA multimers. The disulfide bonds formed by these cysteines in the monomeric forms of IgM and IgA are either interchain (μ and α) or intrachain (α only).

Once interchain disulfide bonds have cemented the four chains into a complete immunoglobulin monomer, the quaternary structure of these molecules can once again be best understood using a domain-by-domain analysis to examine the entire protein. Recall that at this structural level, individual domains interact in such a way that an antibody actually consists of a series of dimeric modules (reviewed in ref. 27). These dimerized domains define the smallest structural and functional units of native immunoglobulin as demonstrated by proteolysis and x-ray diffraction studies. Accordingly, these modules will be reviewed in this context, beginning with the most amino-terminal domain pair, the Fv fragment—or V_H/V_L dimer.

The Fv (see Colorplate 4) is the variable part of the Fab fragment, and as such, constitutes the minimal antigen-binding unit of an antibody. Likely due to this specific functional necessity, V domains dimerize in a manner unlike the strategy employed by all other immunoglobulin domains. In β-pleated sheets, consecutive amino acid side chains protrude on alternating sides at right angles to the plane of the sheet. In most proteins, one side of the sheet packs against another part of the protein (i.e., it is hydrophobic) and the other face of the sheet is exposed to solvent (hydrophilic), leading to a sequence of alternating hydrophobic–hydrophilic residues. In immunoglobulin domains that dimerize, however, this alternating pattern is broken by one of the domain's two β sheets. The sheet that makes up the dimerizing face must interact with both the other β sheet in its own domain and the dimerization face of the adjacent domain. Thus, hydrophilic residues are replaced by hydrophobic side chains that support the dimerization event. While the feature of breaking the alternating hydrophilicity pattern is common to both types of Ig domains that form dimers, V domains interact in a fashion not only specific to V_H and V_L, but also—at
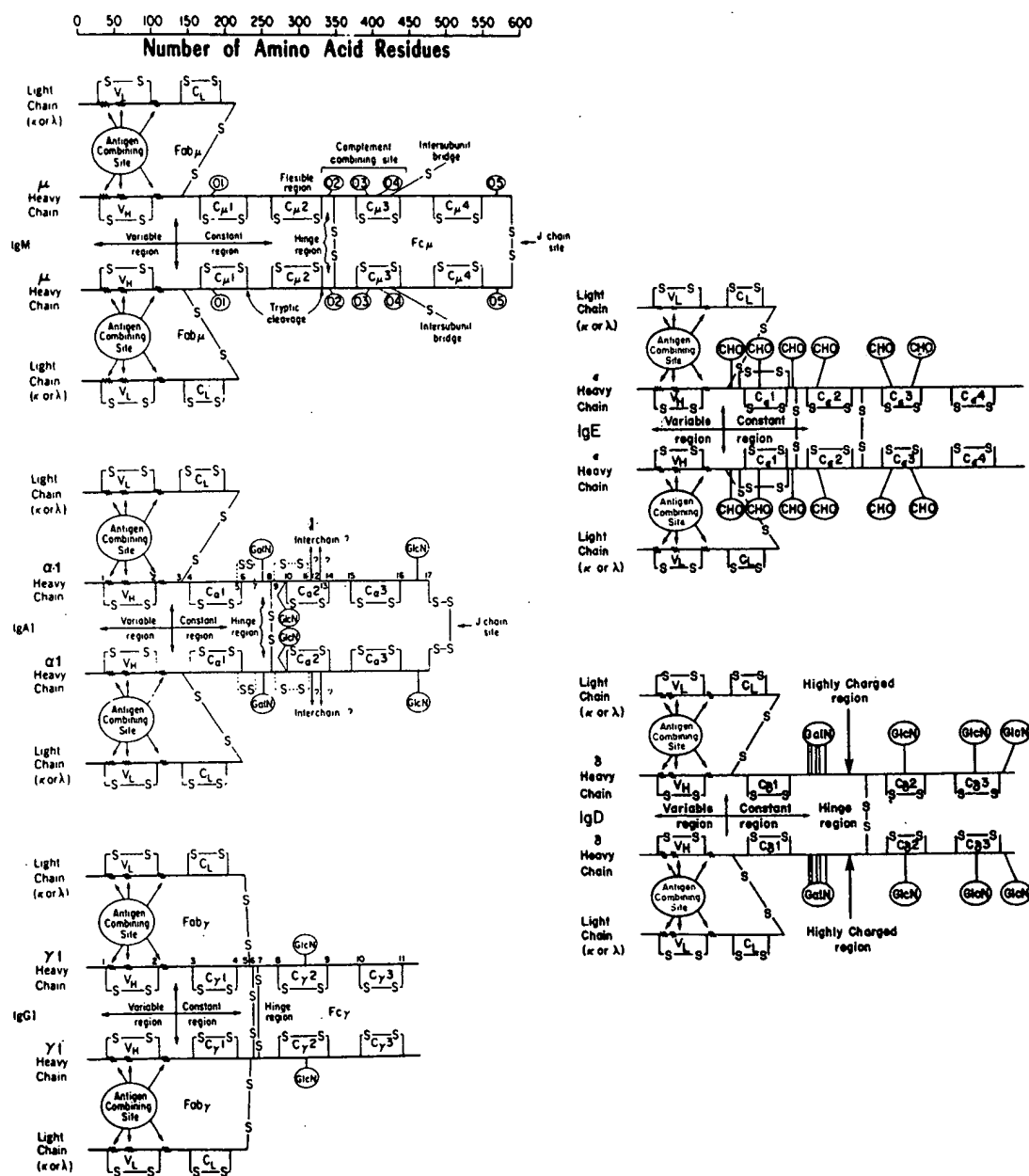
FIG. 16. Schematic representation of the five human immunoglobulin classes. Positions of disulfide bonds and glycosylation are shown for each antibody. (From ref. 66a, with permission.)

the time it was first recognized—unique among all known protein structures (28). Unlike C domains, V domains utilize the five-stranded β-pleated sheet (recall these are unique to V domains) as a dimerization surface (68). The actual V_H/V_L interface consists of the four strands C'-C-F-G on each V domain.

This singular dimerization tactic has profound structural repercussions. In the section detailing tertiary structure, three different β bulges were cited: one common to all V regions found in strand G, one specific to V_H domains located in strand C', and a third unique to V_L also in strand C'. Note that all of these bulges occur in strands at the edges of their respective β sheets. Normally, β sheets pack such that residues in the middle strands form most of

the contacts between layers. On account of these conserved β bulges, however, V domain edge residue side chains protrude into the dimer's interior, preventing a close association of V_H and V_L. The loose packing of the Fv has the effect of creating a hydrophilic groove into which small molecules can fit. This groove, which is primarily lined by residues of the HVRs, and the remaining CDR loops at the end of the V regions form a potential antigen-binding site whose surface area is approximately 2000 to 3000 $Å^2$ (69).

Many of the residues consistently responsible for V_H/V_L interdomain contacts have been localized. About half of the hydrophobic core contacts are formed between FR2 of one chain and FR4 of

the other. The majority of remaining interactions involve CDR3 of one chain and the FR2 and/or CDR3 of the complementary domain. Overall 12 to 21 V_L and 16 to 22 V_H residues participate in interchain stabilization (69). Given the extensive contribution of HVRs to these interactions (28), one might expect that the affinity of H–L pairing might also be variable. Nonetheless, a number of studies using heavy and light chains from different individuals, or even different species, have demonstrated the capacity of heterologous H–L pairs to form stable associations. This implies that the conservation of the basic structural features of V_H and V_L domains has persisted throughout evolution, or at the very least not diverged to the point where the C_H1 and C_L elements are unable to anchor productive interactions between these disparate entities.

Between the Fv and Fb (C_H1/C_L dimer) fragments, a short polypeptide stretch exists that is vital to both the Fv's ability to bind antigen productively and to the ability of C_H1/C_L domains—and all constant domains C-terminal to them—to dimerize properly. This region, comprising the carboxy-terminal amino acids of the V region contiguous with the N-terminal residues of the C region, connects Fv to Fb in the complete Fab fragment and is known as the *elbow peptide*. Collectively, the two elbows of an immunoglobulin Fab are also referred to as the *switch peptide*. Several Fab crystal structures have demonstrated the switch to be a flexible segment permitting considerable bending between the V and C domains (70). This is thought to be important in allowing Fabs to bind epitopes of varying spatial arrangement. An equally important feature of the individual elbow peptides is the fact that they make possible a remarkable 180-degree rotation in the quaternary structure of antibody domains that is essential for the correct orientation of all C domains in order for dimer formation between them (8).

These 180-degree rotations occur at the elbows between V_H-C_H1 and V_L-C_L, and are necessary to properly position the C regions. The most N-terminal C domains, C_H1 and C_L, are then able to combine to form the Fb fragment. C_H1 and C_L are prototypes for the C type Ig domain. Like all C domains, they lack the C' and C" strands present in the V domain five-stranded sheet. Like V domains, they also break the alternating hydrophilicity pattern in one of their two β sheets, but in the case of these domains, this occurs on the four-stranded face of the immunoglobulin fold (A-B-E-D) instead. As a result—and due to the permissive rotations at the elbows—C_H1 and C_L utilize the opposite (relative to V_H and V_L) sides of their domains to dimerize (68). The less-polar residues at the core of these dimeric C modules are conserved in both C_H1 and C_L across species (8–10,12), and in this case (unlike V domains), they tend to reside primarily in the middle strands (B and E), where they mediate a much "tighter" association. Consequently, the Fb is often perceived as a compact anchor for the V domains, forming a stable platform upon which antigen binding can occur. Fb, together with the hinge, also serves as a spacer between the Ag-combining site and the bulky Fc region.

Between the Fab and Fc regions, immunoglobulin hinges (or extra domains) are critical determinants of overall antibody structural and functional properties. Structurally, hinges are extended segments of dimeric peptide held together by one or more disulfide bonds and dominated by prolines, serines, and threonines. This amino acid composition gives hinges flexibility and gives the Fab arms of an antibody flexional (71) and torsional (72) mobility. Hinge flexibility allows the Fabs to conform to the arrangement of epitopes in order to bind bivalently—presumably giving an antibody greater avidity and versatility. The degree of flexing permitted correlates strongly with hinge length between the end of C_H1

and the first interheavy disulfide bridge (the "upper hinge") (71,73); thus IgG3 is more flexible than IgG1. Hinge length and flexibility also reduce the steric barrier that Fabs may present to the access of C_H2 by other molecules. For example, while normal human IgG1 activates complement effectively, the hinge-deleted variant IgG1 paraprotein *Mcg* is unable to fix complement because the Fabs rest too close to the Clq-binding site on C_H2 to allow interaction.

The hinge, being an extended peptide, is the most proteolytically labile part of an immunoglobulin; recall that the early studies that resulted in the understanding of Fab and Fc relied on proteolytic digestion of the hinge. For instance, the δ isotype, with its long, charged hinge, is very susceptible to proteolysis, which may explain its short serum half-life (74). This issue is of critical importance to IgA, which serves its primary function at mucosal surfaces where proteases from bacterial and host sources are prevalent. For example, the α1 hinge contains five carbohydrate attachment sites within a stretch of only 17 amino acids (75), rendering IgA1 resistant to cleavage by intestinal proteolytic enzymes. However, several strains of bacteria secrete proteases that specifically target the α1 hinge. Presumably as an evolutionary consequence, the hinge of IgA2 has undergone a 13-amino acid deletion, restoring its resistance to this second form of proteolytic challenge (53,76,77). Similarly, structural features unique to hinges of each of the different isotypes may have evolved so as to confer their respective immunoglobulins with specific functional characteristics.

C-terminal to the hinge, the Fc region resides. In the cases of IgG, IgA, and IgD, the Fc is a dimer of two C_H2-C_H3; in IgM and IgE it consists of paired C_H2-C_H3-C_H4 domains. Structurally, $C_{\mu/\epsilon}3$ is equivalent to $C_\gamma2$, and $C_{\mu/\epsilon}4$ is homologous to $C_\gamma3$. As has been described above, the vast majority of sites that define the function and physiology of a particular isotype map to the Fc region. The first striking quaternary structural feature of the Fc is that $C_\gamma2$ (and its structural homologues) fail to dimerize. Analysis of these regions demonstrates that C_H2 domains possess a hybrid structure intermediate between V and C domains. Whereas V domains are five-strand/four-strand sandwiches (dimerizing on the five-strand face) and C domains are three-strand/four-strand sandwiches (dimerizing on the four-strand face), $C_\gamma2$ is a four-strand/four-strand sandwich. Moreover, substitutions present in outward-pointing side chains on both sides of the domain prevent dimerization along either face (11). Another mixed feature observed is that the β strands of $C_\gamma2$ are of lengths longer than those of V domains, yet shorter than those of either $C_\gamma1$ or $C_\gamma3$ domains. Finally, all C_H2 domains are derivitized by an N-linked oligosaccharide near the middle of the domain (refer back to Fig. 1), except $C_\alpha2$, where it is found nearer to the carboxy-terminal end of the domain. Hydrogen bonding between these sugars serves as the only contact between C_H2 domains. Moving C-terminally, another important distinction concerning C_H2 is that longitudinal contact with the C_H3 domain (about 340 $Å^2$ surface area/domain) prevents little interdomain bending (12), unlike the flexible elbows between Fv/Fb and the hinge separating Fab and Fc.

The final domains of the Fc, $C_\gamma3$ and its structural equivalents, pair in the manner described for C_H1 and C_L. Studies using limited proteolysis, reduction, and denaturation originally designated this fragment, a dimer of C_H3 domains, as pFc'. C_H3 domains, like the Fb, dimerize with tight association between them (1100 $Å^2$/domain), using the four-stranded faces of each domain (11). Also like C_H1 and C_L, all C_H3-domain isotypes show conservation of core residues involved in this pairwise domain interaction.

Structurally, the only feature that makes a marked distinction between the different C_H3 domain homologues is the presence of the 18-amino acid "tail-piece" that exists at the carboxy-termini of the $C_\alpha3$ and $C_\mu4$ domains. This sequence is important to polymerization and will be discussed further in the section on higher-order immunoglobulin structure. Taken together, the Fc can thus be thought of as an approximation of the Fab, having two pseudo-V regions (the C_H2s) at its amino-terminus, and a module of C-terminal constant regions (the C_H3s) dimerized in the mode characteristic of classic C-type domains.

Although similar in many ways at the protein level, one of the properties that most notably differentiates the quaternary structures of the five classes of Fc is their pattern of glycosylation—with significant functional ramifications. For instance, while the oligosaccharide moiety of the IgG molecule accounts for only 2% to 3% of its mass, it has been shown to be essential for optimal activation of effector mechanisms leading to the clearance and destruction of pathogens (78–80). As introduced above, all human antibody molecules of the IgG class have N-linked oligosaccharide attached at the amide side chain of Asn 297 on the β-4 bend (between β strands D and E) of the inner face of the C_H2 domain of the Fc region (81). This oligosaccharide moiety is of the complex biantennary type, having a hexasaccharide "core" structure (GlcNAc2Man3GlcNAc) and variable outer arm "noncore" sugar residues, such as fucose, bisecting N-acetylglucosamine, galactose, and sialic acid (see Fig. 17). In all, a total of 36 structurally unique oligosaccharide chains may be attached at each Asn 297 residue. It is likely, but not certain, that the precise fidelity of this glycosylation is important.

The site for this C_H2 carbohydrate is a conserved feature for all mammalian IgGs, and glycosylation occurs at a homologous position in human IgM, IgD, and IgE molecules. As stated above, IgA is also glycosylated within its $C_\alpha2$ domain, but at a site further C-terminal. Human IgM, IgA, IgE, and IgD molecules also bear additional N-linked oligosaccharide moieties attached to the C domains of their heavy chains. Furthermore, IgA1 and IgD proteins also possess multiple O-linked sugars in their extended hinge regions, attached to the hydroxyl groups of serine and threonine residues. Glycosylation, in one form or another, is in fact characteristic of all heavy chain C regions and remains one of the most active areas of research in immunoglobulin structural biology.

The structural and functional consequences of Fc oligosaccharides have begun to be assessed experimentally by comparison of

```
                                    (F)
                                     |
(SA) ± (G) ± GN-M                    ±
                        \         M-GN-GN—protein
                         >
(SA) ± (G) ± GN-M       /
```

FIG. 17. Schematic representation of N-linked sugars attached to all C_H2 domains. The core carbohydrate moiety of the complex form of oligandosaccharides is represented by the sugar residues in *open type*. The possible outer-arm residues are *in parenthesis*. All possible combinations are observed. SA, sialic acid; G, galactose; GN, N-acetylglucosamine; M, mannose; F, fucose. Attachment of oliosaccharide occurs on the amide side chain of the Asn-X-Ser/Thr sequon (X ≠ Pro). The Ser/Thr residue forms hydrogen bond(s) with the amide group in order to activate it for attachment to the primary N-acetylglucosamine residue of the dolichol intermediate (catalyzed by oligosaccharyltransferase). (From ref. 81a, with permission.)

glycosylated and aglycosylated forms of IgG. The latter is ordinarily generated by growing IgG-producing *E. coli* in the presence of tunicamycin (a glycosylation inhibitor) or by protein engineering of the carbohydrate acceptor sequence. One characteristic apparently affected by the sugar moieties found on antibodies is the duration of these proteins' existence *in vivo*. Studies of the blood clearance of glycosylated and aglycosylated mouse/human chimeric IgG1 molecules in mice reveal accelerated clearance for the aglycosylated form despite similar half-lives. Additionally, galactosylated and agalactosylated IgG have been investigated to determine the role of outer-arm sugars in complement (C1q)-mediated lysis. The agalactosylated form, produced following exposure to β-galactosidase, has an observed two fold lower activity than the galactosylated form (82). Related studies have substantiated these conclusions (83).

The vital importance of correct glycosylation was further provided by a study using a human/mouse chimeric IgG1 molecule produced in yeast cells, and anticipated to have high mannose-type oligosaccharide attached at Asn 297. The yeast IgG1 product was unable to activate C1q to trigger human complement-mediated lysis of targets, while the same chimeric IgG1 construct expressed in rodent cells (and therefore glycosylated normally) was effective in that regard. A direct role for oligosaccharide in activating the complement cascade is apparent with mannan-binding protein, a lectin that can function as a surrogate C1 component. The specificity of mannan-binding protein is for mannose and N-acetylglucosamine residues, and it has been shown that it can access and bind to terminal N-acetylglucosamine residues exposed on agalactosyl IgG (84).

Studies utilizing the three types of human Fcγ receptors (FcγRI, FcγRII, and FcγRIII) have also attested to the significance of oligosaccharide modifications on antibodies. The IgG subclass specificity of the FcγR suggests that recognition is correlated with the presence or absence of carbohydrate derivatives. This conclusion is supported by the demonstration that aglycosylated human chimeric IgG3 has a reduced interaction with all three Fcγ receptors. Moreover, at the level of function, while haptenated RBCs sensitized with this same aglycosylated IgG3 antibody could still trigger superoxide production by U937 cells, higher levels of sensitization were required compared to normally glycosylated IgG3. The aglycosylated IgG3 also was not recognized by human FcγRII expressed on K562 and Daudi cells, had reduced rosette formation (mediated through FcγRII expressed on human NK cells), and essentially abolished antibody-dependent cellular cytotoxicity (85,86). Clearly these conclusions illustrate that proper Fc glycosylation is—at least in some cases—necessary for normal structural recognition and biologic function of immunoglobulins.

Glycosylation is potentially important outside of the Fc region as well. It has been estimated that up to 30% of polyclonal IgG molecules are also derivitized by oligosaccharide within the Fab region. Since there are no known sites for sugar attachment in $C_\gamma1$ or C_L, this is most likely in the V regions. Of interest in this regard, an analysis of the DNA sequences of 83 functional human germline V_H gene segments revealed five that encoded potential glycosylation sites. Some, but not all, of these are known to be glycolyslated. In one study of protein and cDNA V_H and V_L sequences, about 25% had potential glycosylation sequences, some of which had arisen as a result of somatic mutation and antigenic selection (87). In most circumstances of V region glycosylation studied thus far, the oligosaccharide moiety does not contribute directly to ligand binding, but can exert a subtle influence on protein tertiary and quaternary structure that is essential for full activity of the antibody. Thus,

oligosaccharides occur in many places on immunoglobulin molecules and can affect antibody characteristics as disparate as antigen-binding and the assortment of different Fc-associated functions.

Higher-Order Immunoglobulin Structure—Polymeric Immunoglobulin

One of the most fascinating structural attributes of immunoglobulin is the ability of two classes of heavy chain, IgM and IgA, to form higher order multimeric complexes. IgM and IgA antibodies do not always form polymers, however; monomeric α and μ isotypes exist in forms analogous to those for the γ, δ, and ε isotypes, as well. In addition, polymeric immunoglobulin (pIg) can typically come in a variety of manifestations. The most common forms of these molecules are dimeric (IgA) and pentameric (IgM), although other polymers have also been described. Electron micrographs of murine pentameric and hexameric IgM and human dimeric and trimeric IgA molecules are presented in Fig. 18. Multimerization obviously increases the number of potential antigen-binding sites

FIG. 18. Electron micrographs of immunoglobulin multimers. In (A), a murine IgM pentamer and interpretive diagram (top) and a murine IgM hexamer and diagram (bottom) are displayed. In (B), a human IgA dimer (top) and trimer (bottom) are presented. All magnifications are ×600,000. (From ref. 87a, with permission, and courtesy of K. H. Roux.)

available on the antibody, and this increase in valence translates into enhanced avidity for polymeric, low-affinity epitopes. This is particularly beneficial for antibodies of the μ and α classes, which serve as the first line of defense at mucosal surfaces where encounter with this type of pathogenic target (i.e., cell-wall polysaccharides) is most frequent. Moreover, IgM, which is the antibody characteristic of primary humoral responses when affinity maturation has not yet occurred, is reliant upon this increased avidity to mediate its functional responsibilities. In addition to raising the apparent affinity for antigen binding, polymeric IgM's juxtaposition of several Fc regions in close proximity also likely plays a role in its efficacy in fixing complement components of the classical pathway. Mechanistically, assembly and secretion of pIg involves the covalent linkage of concatomers of prototypic immunoglobulin monomers, and two accessory proteins termed *J* (joining) *chain* and *secretory component* (SC) play key roles in these processes.

J chain is a 137-amino acid polypeptide synthesized by pIg-producing plasma cells. J chain covalently interacts with one or more cysteines of immunoglobulin monomers undergoing multimerization (88). It is a proteolytically labile molecule with a high content of negatively charged amino acids and has eight cysteine residues involved in both intra- and interchain disulfide bonds (89). The high level of conservation between J chains of human (90), murine (91), rabbit (92), and even amphibian (93) origin implies that there is a powerful selective pressure to maintain J chain structure. A report identifying J chains in invertebrates which have no known correlate to antibody proteins (94) also indicates that the J chain probably performs some basic protein function that predates its eventual development of the ability to interact with immunoglobulin. In any event, structural studies imply that despite the fact that J chain lacks any significant sequence homology with immunoglobulin, it likely folds into a β-barrel structure similar to that of an immunoglobulin fold (95). Besides the intrachain cystines which stabilize the J chain itself, additional cysteine residues form disulfide bridges to the tailpiece of one or more immunoglobulin monomers during multimer assembly (89). Although the actual details of polymerization have not as yet been elucidated, it is known that the 18-amino acid $C_{\mu/\alpha}$ tailpiece and its penultimate cysteine residue are necessary for the process (52,96).

The stoichiometry of multimer assembly is such that one J chain is present per polymer (whether dimer, trimer, pentamer, etc.). While not always a part of pIgM molecules, J chain is probably absolutely necessary for formation of polymeric IgA, as it is always present in the complex. J chain synthesis is known to be highly regulated (97), and it is thought to be linked to the B cell's activation state as well (98). High levels of J chain expression have been shown to result in production of normal J chain–associated pentameric IgM, while low J chain synthesis results in secretion of hexameric IgM lacking the protein. Intriguingly, this hexameric IgM is actually 20-fold more potent at promoting lysis by complement than is the usual pentamer (99).

The second accessory molecule associated with the secretion of multimeric antibodies actually derives from another protein belonging to the IgSF—the pIg receptor (pIgR)—and is not even made by cells of the B lineage. Secretory component (also called *secretory piece*) was initially discovered as a polypeptide found tightly complexed to the Fc of secreted forms of IgA and IgM (100); subsequently, it was recognized that SC is actually a portion of the larger transmembrane pIgR protein (101). The entire cDNA sequence of the pIgR reveals a protein consisting of seven domains: the first five are extracellular and structurally similar to
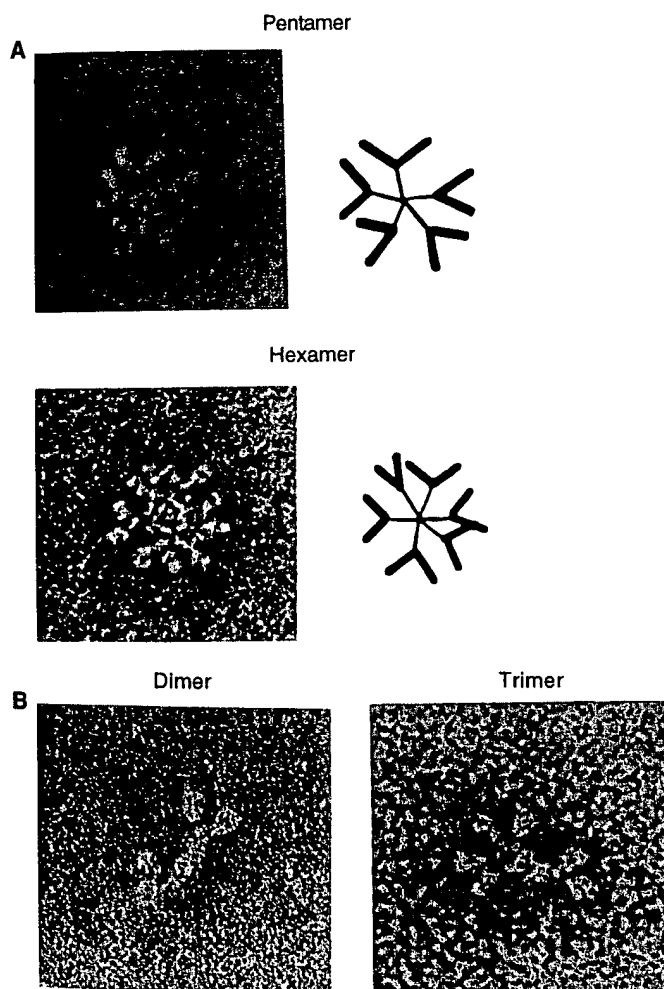
immunoglobulin V regions, the sixth contains a transmembrane segment and is partially homologous to immunoglobulin V domains, and the seventh contains an unrelated C-terminal intracellular domain (102). The first five domains of the pIgR are in fact the secretory piece originally co-isolated as part of the secreted immunoglobulin complex.

The pIgR is synthesized in epithelial cells of the respiratory, gastrointestinal, and genitourinary tracts and is expressed on their basolateral aspect, where it binds to pIgA and pIgM in a high-affinity interaction. It is known that the N-terminal domain of the pIgR confers binding specificity, and it is thought that both J chain and Fc $C_\alpha 3/C_\mu 4$ determinants are recognized by the receptor. Interestingly, although the precise molecular locations of pIg/SC interaction have not been identified, there is evidence that, at least structurally, the site is well conserved. Studies have shown, for instance, that human SC binds not only human pIgA and pIgM (103,104), but also several other mammalian species' IgM and IgA (105), and even chicken IgA (106). This cross-reactivity, however, may be mediated by the J chain rather than the Fc regions.

Regardless, following the initial interaction between domain 1 of the pIgR and the C-terminal domain/J chain of the pIg, secondary contact occurs between pIgR domains 3, 4, and 5 with the antibody, consumated by formation of a disulfide linkage between the SC and $C_\alpha 2/C_\mu 3$. This covalent bond is between Cys 467 in domain 5 of the pIgR and Cys 311 located in the $C_\alpha 2$ domain of one IgA subunit's heavy chain (an IgA dimer would have four $C_\alpha 2$ domains overall) (107). After a stable interaction has been established, endocytosis of the complex occurs via clathrin-coated pits. Next, following cleavage between domains 5 and 6 of the pIgR, the poly-Ig /SC (now formally termed *secretory immunoglobulin*) is exocytosed at the apex of the cell, releasing the secretory immunoglobulin onto the mucosal surface (108). Current thinking holds that SC's most important function, outside the realm of its role as the pIg-binding portion of the pIgR, is to help protect secretory immunoglobulin in harsh mucosal environments.

Structurally then, polymeric antibodies represent the pinnacle of complexity in terms of immunoglobulin's expansion upon the Ig homology domain concept. From the fundamentals of a simple 110-amino acid domain with a few conserved core residues and a basic structural topology, an intricate molecule such as pentameric IgM (containing 70 different Ig domains of both V and C types, not to mention Ig domains contributed by SC!) is constructed. A molecule capable of recognizing as many as ten (although steric constraints usually dictate less) identical specific antigenic determinants, and also able to mediate several different important biologic functions—all of which will be detailed in the following section.

IMMUNOGLOBULIN FUNCTION

Throughout this chapter thus far, many of the differing functional capacities of antibodies have already been alluded to, as pertains to the identification of the specific structural determinants responsible for particular interactions. Still, the plethora of biologic activities performed by immunoglobulin is best treated as a separate section, in which the many and varied aspects of immunoglobulin function can be detailed and integrated in a physiologic context. Collectively, secreted antibodies are able to activate both the classical and alternative complement cascades (see Chapter 29), transcytose across epithelial cell layers to provide a

barrier to pathogens at mucosal surfaces (see Chapter 27), travel transplacentally to confer maternal humoral immunity to the fetus and neonate, induce phagocytosis by macrophages and granulocytes via the process of opsonization (see Chapters 30 and 41), foster antibody-dependent cellular cytotoxicity by lymphocytes and NK cells (see Chapters 17 and 31), encourage antiparasite immune responses by eosinophils (see Chapter 38), and promote degranulation by mast cells and basophils (see Chapters 32 and 35)—not to mention antibody's ability to bind and inactivate foreign antigenic entities directly (see Chapter 39)!

Even this imposing list of attributes neglects to mention the myriad effects mediated by surface immunoglobulin that include, but are not limited to, the induction(s) of activation (see Chapter 7), differentiation (see Chapter 6), anergy (see Chapter 20), and even apotosis (see Chapter 23) of B lymphocytes, which are detailed elsewhere in this volume. Surface Ig on memory B cells also has the ability to act as a high-affinity receptor for the recognition, internalization, degradation, and eventual presentation of specific antigens to T cells (see Chapter 9). This allows memory B cells to act as *antigen-specific* antigen-presenting cells (APC), which makes them uniquely efficient among this class of cells. Moreover, emerging fields of study, such as the growing body of literature concerning intracellular antibodies, indicate that new functional capacities for immunoglobulin are likely yet to be discovered.

As for the preceding sections on immunoglobulin structure, the biologic capabilities of immunoglobulin are best treated by dissecting the molecule into the portions responsible for each of its different functional characteristics. While there are some exceptions, in general, specific functions of antibodies can be ascribed to individual domains of the molecule. In the case of variable regions, this requires consideration of the two V domains (V_H and V_L) whose primary function is the binding of antigen. Additionally, it has also been appreciated that certain "superantigens" bind to the V domains as well. In the case of constant regions, because no effector properties have been linked to C_L domains, this entails discussion of each of the heavy chain isotypes (IgM, IgD, IgG, IgA, and IgE), whose functional differences must be a direct result of their structural heterogeneity.

Variable Region Functions

The two V regions (either V_H/V_λ or V_H/V_κ) together form the variable domains of an antibody molecule and provide the specificity for targeting the effector arms of immune response. In general, both V regions are needed to provide specificity and high affinity. There are many examples of individual variable regions binding antigen, but clearly, when the two chains act in concert, the exquisite specificity and affinity of interaction between antibody and foreign antogen is dramatically enhanced.

The concept of hypervariable regions and complementarity-determining regions is pertinent here. In general, most of the contacts between the V domains and antigen take place between amino acid residues in the CDRs and the major epitopes on the antigen (109). However, more recent studies have documented considerable contact between so-called framework residues and antigen. This is most dramatically seen in the lysozyme–antibody crystal (see Colorplate 6) where many residues (especially in the heavy chain FR3 region) are in direct contact with the antigen (110). The generalization that the CDRs provide *all* of the contact residues grew out of the early work involving hapten/anti-hapten systems. When only

small organic haptens are the "antigen," then the CDRs can easily provide a "pocket" in which antigen engages antibody (recall that the peculiar dimerization strategy employed by V domains has the propensity to generate such pockets). Closer inspection of the lysozyme Ag-Ab complex—in particular, lysozyme residue Gln 121—is instructive in this regard. Gln 121 protrudes into the cleft between V_H and V_L much like haptens fit into the groove described above. However, other non-groove residues still appear to provide the bulk of the interacting surface for this antibody. In fact, other antibodies that are also reactive with lysozyme (see Colorplate 6) have further borne out this face-to-face binding concept (59). It is fair to say, then, that generally when large molecules such as proteins complex with antibody, the interaction is one of two protein "faces" coming together. In that instance, the notion of a pocket is less appropriate, and as such, non-CDR residues (especially in FR1 and FR3) are also frequently involved.

As with all molecular associations, antigen–antibody interactions occur only if the binding reaction releases enough free energy to be thermodynamically favored. The affinity of interaction is exponentially related to changes in free energy (see Chapter 4). Free energy changes are the sum of changes in both entropy and enthalpy, with increases in entropy and decreases in enthalpy favoring binding. Few association reactions are able to fulfill both of these requirements, however. Instead, a favorable change in one component compensates for a less unfavorable change in the other. When antibodies bind their ligands, the freedom of one molecule to move relative to the other is lost (an unfavorable decrease in entropy). Stabilization of most conformational motions of both the epitope and the backbone and side chains of the paratope surface lowers entropy even further. Thus, to encourage binding to antigen, antibodies must attempt to limit decreases in entropy and offset these losses by potentiating decreases in enthalpy. At the amino acid level, this leads to a selection for Tyr, Trp, Ser, and Asn in combining sites (111,112), because these residues have lower conformational freedom, and hence less entropy to lose upon binding. Additionally, the side chains of these residues foster the varied chemical interactions that drive changes in enthalpy necessary to promote binding energetically.

Specifically, the antigen–antibody interaction involves a variety of forces, including electrostatic (the attraction between opposite charges), hydrogen bonds (hydrogen shared between electronegative atoms), van der Waals forces (the fluctuations in electron clouds around molecules oppositely polarize neighboring atoms), and hydrophobic forces (hydrophobic groups interact unfavorably with water and tend to pack together to exclude water molecules) (113). Of course, salt bridges and other forms of interaction also come into play in some specific immunoglobulin–ligand complexes as well. It is also important to appreciate that rarely do covalent bonds occur between antigen and antibody. Thus, antigen–antibody complexes are readily dissociated by solvents that break the above bonds, such as high salt, organic solvents, urea, and so on.

Thermodynamic considerations for ligand binding are favored by large interacting surfaces of both antibody and antigen, which are packed as closely as possible. Large interaction areas exclude more bound water, somewhat opposing losses in protein entropies with gains in solvent entropy. Surprisingly then, some antigen–antibody complexes actually retain water molecules in their interfaces. However, rather than interfering with binding, these frequently contribute to the interaction by hydrogen bonding to both surfaces (114). More important than entropic changes, the overriding impetus for large contact areas of antibody and antigen is their

ability to bring about large decreases in enthalpy. This effect is mediated by allowing many chemical interactions of all kinds to occur simultaneously between the epitope and paratope.

A model for interaction outside the realm of typical antibody–antigen binding has recently come into vogue, that of the "superantigen." Superantigens were first appreciated in the context of T cell activation. Certain molecules (particularly bacterial products) were found to interact with many different T cell receptors (TCRs) having a variety of specificities (115). Thus, superantigens were originally defined as intact proteins that stimulated large numbers of T cells by binding the V region of a specific family of V_β chains (the heavy chain of the TCR) outside its normal binding groove. Typical T cell superantigens simultaneously stimulate 5% to 25% of the T lymphocyte population, compared with 0.01% stimulation of T cells by a conventional antigen.

More recently, this concept has been extended to B cells (116,117). SPA is a prototype of a B cell superantigen. Although SPA was known to bind to certain immunoglobulin C regions—it has long been used as a mitogen for human B cells—it has been shown that SPA also binds to certain human V_H3-encoded antibodies (118,119). It also binds to the Fab region of murine immunoglobulins, particularly those of the J606 and S107 V_H gene families (which belong to the same clan as human V_H3). SPA binds independently of D, J_H, and light chain utilization, although some light chains influence the extent of binding. Since the interaction is independent of the specificity of the antibody, and since SPA in general does not block antigen binding, it is considered a B cell superantigen. SPA is even able to deliver activation signals to stimulate the differentiation of those B cells containing V_H3-encoded receptors, and SPA also stimulates antibody production. More recent work (120) has documented that SPA simultaneously interacts with FR1, CDR2, and FR3 on the V_H region (Fig. 19).

Other superantigens have also been described that are able to bind immunoglobulin in regions apart from the traditional antigen-
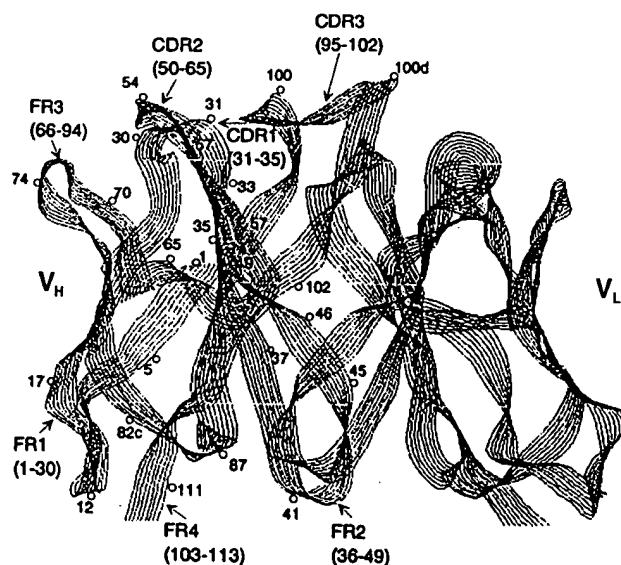
FIG. 19. Ribbon drawing of the Fv fragment of the V_H3 antibody KOL. The FR1, CDR2, and FR3 subdomains of the heavy chain (left) are juxtaposed in a manner forming a solvent-exposed face which allows SPA binding. (From ref. 120, with permission.)

combining site (based on their broad specificities). These include the HIV envelope protein gp120 (121), which like SPA binds V_H3-encoded antibodies, and the TCR-associated molecule CD4 (122). Like the circumstance with antigen binding, superantigens generally require both V_H and V_L domains (even though the particular identity of the light chain is unimportant). Individual heavy chains do not bind the B cell superantigens that have been described to date, indicating that light chains must influence their conformations appreciably. Finally, the report of a crystal structure (an IgM rheumatoid factor Fab complexed to its autoantigen, an IgG Fc) showing residues at the edge of the conventional binding site mediating interaction indicates that still more novel paradigms for antibody–antigen binding possibly exist as well (123).

Constant Region Functions

Because mammalian species each utilize the same major classes of antibody (although their organization of subclasses differs), it is reasonable to presume that each isotype subserves some vital biologic function(s). Along these same lines, it should be remarked that even in "lower" species, where only one type or one copy of heavy chain gene is present, the protein product resulting from this element is typically heterogeneous. In other words, although fish, amphibians, and reptiles all possess fewer immunoglobulin isotypes than do mammals at the genomic level, greater than one C region protein is produced per gene. For example, sharks make both monomeric and polymeric IgM; in skate, turtle, and duck there are truncated and full-length versions of the immunoglobulin polypeptide; Xenopus immunoglobulin comes in both glycosylated and aglycosylated forms (51). Clearly, evolution has recurrently employed the strategy of adopting more than one type of antibody

to perform the multitude of biologic responsibilities that are required by species for effective immunologic functioning.

In a broad sense, Fc-mediated effector functions can be classified into three general categories: (a) activation of complement, (b) interaction with effector cells, and (c) transport and compartmentalization of immunoglobulins. In addition, different isotypes have different stabilities in vivo, such that this is an important variable as well. In the following sections, the five classes of human immunoglobulins are each discussed separately with respect to function. Table 2 presents a summary of key properties for each class of human antibody, and Fig. 20 compares the circulating serum levels for each of the five major isotypes.

IgM

IgM is the most versatile antibody and almost certainly the first type of immunoglobulin to have developed evolutionarily. Heavy chains of the μ class are the first type expressed during B cell development, and IgM is the isotype produced in primary immune responses. IgM, in the form of surface immunoglobulin, is also an important receptor on immature B lymphocytes and on mature, naive B lymphocytes. Total serum Ig consists of 5% to 10% IgM, and second to IgA it is the major isotype of mucosal immunity. Originally named due to their description as macroglobulins, IgM molecules are thought to serve similar functions in all mammalian species. In fact, IgM-like (polymeric, having five domain heavy chains with large carbohydrate content, and present as a cell surface receptor on most B cells) antibodies have even been identified in most non-mammalian vertebrates other than the jawless fish (51). Unquestionably, the polymeric structure of IgM has been conserved in evolution, probably due to its higher avidity for antigen compared with that of the monomer.

TABLE 2. *Physical, chemical, and biological properties of human heavy chain immunoglobulin classes*

Property	IgM	IgD	IgG	IgA	IgE
Molecular form	Pentamer, hexamer	Monomer	Monomer	Monomer, dimer	Monomer
Number of C region domains	4	3	3	3	4
Tailpiece	+	–	–	+	–
Accessory chains	J chain, SC	None	None	J chain, SC	None
Subclasses	None	None	G1,G2,G3,G4	A1,A2	None
Molecular weight	950 kD, 1150 kD	175 kD	150 kD	160 kD, 400 kD	190 kD
Carbohydrate content (%)	10	9	3	7	13
Percentage of total serum Ig	5–10%	0.3%	75–85%	7–15%	0.02%
Average adult free serum level (mg/ml)	0.7–1.7	0.04	9.5–12.5	1.5–2.6	0.0003
Synthesis rate (mg/kg/d)	7	0.4	33	65	0.016
Serum half-life (d)	5	3	23	6	2.5
Antibody valence	10, 12	2	2	2, 4	2
Bacterial lysis	+	?	+	+++	?
Placental transfer	–	–	+	–	–
Mast cell/basophil binding	–	–	–	–	+
Macrophage binding	–	–	+	+	–
Classical complement activation	++	–	+	–	–
Alternate complement activation	–	+	+	A1+,A2–	–
Other biological properties	Primary Ab responses; Secretory immunoglobulin	Unknown; Useful as a B cell marker	Hallmark of secondary immune responses	Main secretory immunoglobulin	Allergic and anti-parasite responses
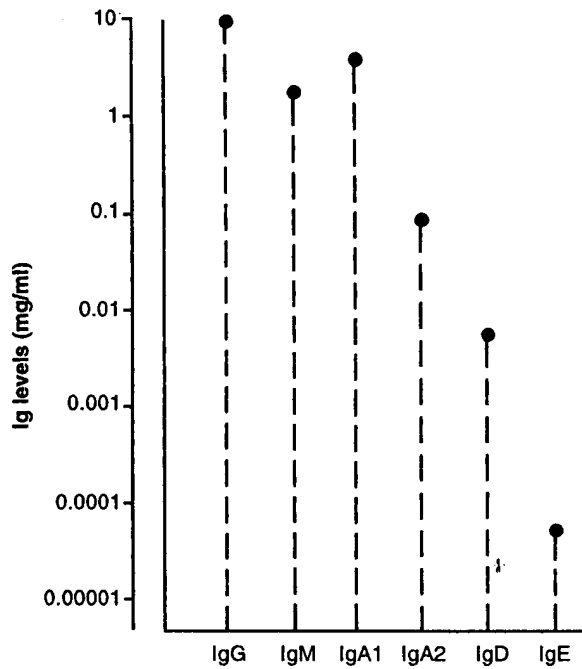
FIG. 20. Circulating levels of different human immunoglobulin iso-types. Note the log scale of the graph and that both human IgA iso-types are represented. (From ref. 123a, with permission.).

The two most common forms of IgM are the membrane-bound monomeric form and the secreted pentamer. The cell-surface version of IgM serves as the antigen-specific receptor for B cell activation, although the activation signal is actually transmitted by the transmembrane accessory molecules Igα and Igβ (124). It is unclear whether the $C_\mu 4$ domains of surface IgM participate in the interaction with the α/β heterodimer, but it is more likely that the important associations lie in the transmembrane and cytoplasmic regions of IgM that are specific to the cell-surface form (125). The surface form of IgM is also important in the development of the B cell. During pre-B stages, μ heavy chains are associated (via a disulfide linkage in $C_\mu 1$) with the "surrogate" light chains V_{preB} (analogous to a V_L domain) and $\lambda 5$ (a C_λ analogue) (126,127). This complex, once again through accessory molecules, is able to transduce signals thought to be necessary for allelic exclusion of the other heavy chain locus and for subsequent light chain rearrangement (detailed in Chapter 5). Eventually, of course, the IgM heavy chains become associated with either λ or κ light chains.

Polymeric IgM also has its own catalogue of functional attributes. IgM antibodies are the first to be secreted from plasma cells upon challenge by antigen; since IgM is not secreted in large quantities from memory B cells, elevated IgM is indicative of recent antigenic exposure. As stated earlier, IgM antibodies generally have low affinity, as they have not gone through the processes of somatic hypermutation and affinity selection. Nonetheless, the high avidity of pIgM renders it capable of efficiently binding antigen. Similarly, a single polymeric IgM molecule is able to effectively initiate classical complement fixation, even though the affinity of C1q for $C\mu$ is very low. The C1q-binding site of IgM has been localized to the $C_\mu 3$ domain (128) and appears to be dependent upon carbohydrate found there for potent binding (129). While $C_\mu 3$ domains (and their structural homologues) are not well conserved evolutionarily, the ability

to mix IgM and complement from different species and retain activity indicates that the complement recognition sites on vertebrate immunoglobulins may be similar. In fact, even *Xenopus* IgM has been demonstrated to fix mammalian complement components (130)! IgM has also been shown to interact with C3b via its $C_\mu 1$ domain, thereby allowing antibody–antigen complexes containing IgM to indirectly mediate phagocytosis. By this mechanism, C3b, once fixed, can promote uptake via complement receptors found on macrophages.

The high avidity of IgM for both antigen and complement is crucial in the context of its role as a front-line defense mechanism. IgM not only is the humoral agent of primary immune responses, but also—like IgA—is transported by the pIgR across epithelia such that it serves a role as a secretory immunoglobulin at mucosal surfaces. Since secretory immunoglobulins are present in breast milk as well, IgM also participates significantly in protecting the newborn from intestinal pathogens until such time as the neonatal immune system is fully functioning. A role for IgM in mucosal immunity must have developed early in evolution, as it is the sole immunoglobulin in some animals.

IgD

IgD is present in serum in very low amounts (less than 0.5% of total serum Ig). Although synthesis of IgD is also very low (at least an order of magnitude lower than that of IgM, IgG, and IgA), IgD's pronounced susceptibility to proteolytic degradation is probably also responsible for its scarcity in plasma and other bodily fluids. The unusually long hinge region linking Fab to Fc in IgD is thought to be largely accountable for its short half-life. IgD is secreted neither during an immune response, nor following mitogenic stimulation of IgD+ B cells, although in the form of immune complexes it is known to be able to activate the alternative complement cascade. IgD's low levels make this complement fixation unlikely to be important in the *in vivo* context. In fact, no specific functions unique to IgD have been definitively assigned to the δ Fc region in either its membrane-bound or soluble forms. That notwithstanding, the fact that the IgD class is maintained in all mammals, has a high level of conservation across species (131), and the existence of an Fcδ receptor, all suggest that it may have some distinct purpose. Still, two independently derived strains of IgD knockout mice have failed to ascribe to it a convincing immunologic role. In one strain, in which a premature stop codon was introduced into the $C_\delta 3$ domain, a subtle reduction in the total number of peripheral B cells was noted (132). In the other, which carried an insertion in its $C_\delta 1$ exon and a frameshift in $C_\delta 3$, delayed affinity maturation during T cell-dependent antigen responses was demonstrated (133).

Although not known to have any unique function, IgD, together with IgM, is a major surface component on many B cells. Because the C region genes for μ and δ are both transcribed in the same primary RNA message, differential splicing to produce either IgM or IgD is required. This particular genomic organization facilitates their coexpression, which is not possible for any other isotypes (reviewed in Chapter 5). Mature, naive B cells migrate from the bone marrow as IgM+/IgD+ cells (134) and make up about 90% of peripheral B cells in both the murine and human systems (see Chapter 6). Similarly, B cells in the primary follicles of secondary lymphoid organs coexpress IgM and IgD, but as they mature to memory cells, IgD expression is typically lost (135,136). Curiously, studies of IgM+/IgD+ splenic B cells reflect that IgD surface

expression is actually tenfold higher than IgM levels (137). This is particularly puzzling, given that δ message levels are lower than are μ mRNAs, and IgD (at least in the serum) is known to be so proteolytically labile. Perhaps helping to explain this high level of IgD expression is that fact that IgD does not need to complex with other proteins for transport to the cell surface, distinguishing it from all other immunoglobulin classes (138). It is possible that IgD's high levels of surface expression and intrinsic flexibility (139) afford it a role in the early response to antigen (123a,140).

In addition to their coexpression, IgM and IgD have a number of commonalties in terms of their function as B cell antigen receptors. Like IgM, IgD is non-covalently associated with Igα/Igβ heterodimers, which serve as the signaling component of their BCR (138). Not surprisingly then, ligation of either IgM or IgD by antigen can independently mediate activation, deletion, or anergy of B cells (141), and likewise, the signals propagated by IgM or IgD BCR seem to be the same, albeit with different kinetics (see Chapter 7). Specifically, signals transmitted through surface IgD have been reported to cause induction of APC function (142); upregulation of coreceptors B7-1 and B7-2 (143); class switching to IgM, IgG1, IgG2, IgG3, and IgA (144); and increased secretion of IgE (145). The biologic significance of many of these findings remains unclear. In fact, reports (146,147) describing a new class of germinal center IgD+ B cells (having evidence of up to 80 different somatic mutations within their V regions!) demonstrate just how little is still understood about the cells expressing—and the protein—IgD.

IgG

IgG is the predominant immunoglobulin in blood, lymph, peritoneal fluid, and cerebrospinal fluid. Collectively, it makes up more than 75% of serum immunoglobulin and is synthesized at a high rate (over 30 mg/kg/d, second only to IgA). The presence of high-affinity IgG is the hallmark of secondary humoral immune responses. Electrophoretically, IgG proteins migrate to the γ range of serum globulins, hence IgG's earlier designation as gamma-globulin. Actually, IgG is composed of four subclasses of antibody, whose salient features are summarized in Table 3. The selection of IgG subclass by a particular immune response does not appear to be random: in murine systems, anti-carbohydrate specificities tend to be IgG3, anti-protein IgG1, and anti-viral IgG2a (148,149). In man, reactivities against polysaccharide immunogens are skewed toward IgG1 and IgG2, while anti-protein and anti-viral γ antibodies are biased in the direction of IgG1, IgG3, and IgG4 (150). As

an offshoot of these phenomena, clinical syndromes in which specific IgG subclasses are absent are known to present themselves as characteristic immunodeficiencies (see Chapter 43).

Perhaps the most studied feature of the IgG isotypes is their ability to activate the classical complement pathway. Although all four are capable of initiating the classical cascade, they do so to varying degrees (G3>G1>G2>G4) (150,151). Understanding the means by which the different IgG subclasses interact with specific components of complement has been difficult, complicated by many confounding reports. Results indicating that C1q is unable to bind either IgG2 or IgG4 antibodies (152) were perplexing, given that both are able to activate the classical cascade. Similarly, despite Fcγ3's higher affinity for C1q (152), IgG1 is more effective at potentiating complement-mediated cytolysis. When the site for C1q-binding was mapped to the C-terminal portion of the $C_\gamma2$ domain near the hinge region (153,154), investigators surmised that differences in the IgG subclasses' abilities to activate complement were likely attributable to steric freedom, or lack thereof, conferred by the particular hinge of the antibody (155,156). For this reason, the longer hinge of the γ3 C region was thought to account for IgG3's higher affinity for C1q, relative to that of IgG1. Still, hinge-deletion (157) and hinge-swapping (158) experiments have yielded data that contradict the notion of the hinge being a key determinant for complement activation. Be that as it may, recall that proper glycosylation within the $C_\gamma2$ domain is accepted as an obligatory element for fixation of complement as well.

The explanation for the difference in efficacy of lysis by complement between IgG1 and IgG3 (paradoxical, given their affinities for C1q) is even more convoluted. It may reflect that other differences between the IgG1 and IgG3 C1q sites are present that affect complement activation. Alternatively, it may derive from differences in a second, separate site in the $C_\gamma1$ domains of these molecules which, like IgM, binds activated C3b and protects it from inhibition. This second site also likely explains the capacity of IgG2 and IgG4 to activate the classical pathway, despite an inability to bind C1q. Finally, note that IgG4 is able to efficiently recruit and activate the alternative complement cascade, distinguishing it from the other three subclasses.

Another means by which IgG antibodies communicate with the effector arms of the immune system is via the Fcγ receptors (FcγRs). A number of different IgG FcR exist (covered specifically in the section on the IgSF), each of which have their own profile and affinities for binding of the different IgG subclasses, expression patterns on different cell types, and different biologic responsibilities (159). Among the immunologic cell types implicated as

TABLE 3. *Properties of Human IgG subclasses*

Property	IgG1	IgG2	IgG3	IgG4
Disulfide linkages	2	4	5–15	2
Molecular weight	146 kD	146 kD	165 kD	146 kD
Percentage of total serum immunoglobulin	34–87%	5–56%	0.5–12%	7–12%
Average adult free serum level (mg/ml)	5.9 ± 2.6	3 ± 2.5	0.6 ± 0.55	0.9 ± 0.25
Macrophage binding by FcγR	+	–	+	+
Placental transfer	+	++	+	++
ADCC	+++	+	+++	+
Classical complement activation	+++	++	++++	+
Alternative complement activation	+	+	+	+++

Adapted from Simard and Mak (140).

important binders of IgG are macrophages, polymononuclear cells, and lymphocytes (including B cells). Interactions with these receptors cause many functional effects, including phagocytosis (160) and antibody-dependent cell-mediated cytotoxicity (161), both of which ultimately lead to the destruction of the bound antigen. Specifically, the hierarchy for ADCC by mononuclear cells is IgG1, IgG3 > IgG2, IgG4 (152,162). Signals transmitted via FcγR also modulate lymphocyte function by means of up-regulation or down-regulation of antigen presentation, cytokine release, cytokine receptor expression and/or sensitivity, and even immunoglobulin secretion. Even soluble FcγR are known to bind IgG, although the significance of this finding is unclear (163). Finally, IgG FcR also permit transplacental movement of maternal antibodies during gestation (164). This provides the developing fetus with a source of high-affinity serum immunoglobulin that is able to interact with complement to mediate biologic effects at a time at which it has no other form of specific humoral immunity. It should not be overlooked that IgG molecules are the most stable isotype in serum (with a half-life of over 3 weeks), further maximizing their utility in this endeavor—even into the post-natal period.

Binding of the four IgG subclasses by the different FcγR varies in terms of the specific contact residues involved for each respective ligand–receptor pair. Generally, although the IgG binding sites for the FcR are thought to largely overlap, the precise elements responsible for interaction likely have subtle differences. By consensus, research into these issues has suggested that the sites are bipartite, consisting of a site on the C-terminal portion of the hinge and also reliant upon residues found in the portion of the $C_\gamma2$ domain already implicated in C1q-binding (86,165). Because the four IgG isotypes differ considerably in these regions, this would fit nicely with their noted differential binding of the varied FcγR.

IgA

IgA is the major immunoglobulin in external secretions such as saliva, mucus, sweat, gastric fluid, and tears. Moreover, it is also the major immunoglobulin of colostrum and breast milk, where it provides the neonate with a readily available source of intestinal protection against pathogens (167). The secretory forms of IgA are exclusively polymeric, including J chain and SC in the manner described previously. In addition, IgA—present predominantly in its monomeric form—is also an important component of serum Ig, where it makes up 10% to 15% of the total. The synthetic rate of IgA is roughly double that of IgG, such that total daily IgA production outpaces that of all other immunoglobulins combined. The majority of IgA synthesized is in the secretory form, with the largest fraction of IgA plasma cells residing in the subepithelial mucosa of the small intestine. Because secretory IgA coats all external surfaces except skin, it is rightly considered a first line of defense against organisms that would invade via mucosal routes. IgA's role in mucosal immunity (see Chapter 27) is phenotypically evident in persons with the most common genetic defect of the humoral immune system, IgA deficiency (see Chapter 43). Individuals with this condition are susceptible to invasion across mucosal barriers and typically present clinically with recurrent infections of this type.

Serum and secreted IgA originate from separate pools of B lymphocytes: plasma cells in specialized sites of the respiratory, urogenital, gastrointestinal, and mammary tissues produce the IgA found in secretions, while the IgA in serum emanates from plasma cells in the bone marrow, lymph nodes, and spleen. Despite this compartmentalization of production, antigenic exposure occurring at either mucosal or systemic sites will prime the development of both secretory and serum IgA responses simultaneously (168).

In humans, the two IgA subclasses, IgA1 and IgA2, show an interesting division of expression which affects their resultant biologic utilities. IgA1 exists primarily as a monomeric molecule, and accordingly is the main IgA isotype in plasma (refer to Fig. 20). In bone marrow, about 90% of IgA-secreting plasma cells make IgA1 (169). IgA2, on the other hand, is usually found as a polymer. Recall that the main structural difference between these two isotypes is localized to the hinge. Whereas the IgA1 subclass has a higher concentration of carbohydrate in its hinge region (protecting it from most forms of proteolytic degradation), the IgA2 isotype has deleted much of that hinge region—presumably as an evolutionary response to bacterial IgA1-specific proteases (170). Thus, it is consistent that the broadly protease-resistant form (IgA1) should predominate in serum to maximize its lifespan, while the targeted protease-resistant subclass (IgA2) should prevail where bacterial exposure is more common.

IgA does not efficiently induce inflammatory responses. Rather, it is believed to protect primarily by exclusion, binding and cross-linking pathogens to prevent their uptake across epithelia and facilitating their expulsion in mucus excretions (123a,140). It is noteworthy that inflammatory responses localized to mucosa would likely be detrimental to barrier function, as tissue damage could compromise the integrity of epithelial surfaces. While IgA does have the ability to fix complement via the alternate cascade, this ability is restricted to IgA1. IgA can also opsonize antigens for phagocytosis; this is accomplished via a specific Fcα receptor (FcαR) found on macrophages, monocytes, and neutrophils. This provides a mechanism for IgA immune complexes that accumulate at mucosal surfaces to be engulfed and processed. The FcαR is known to bind secretory IgA with higher affinity than serum IgA, but strangely, the site on IgA that is recognized seems to be unrelated to the J chain or SC which distinguish the secretory and serum forms (66). Rather, in a manner unique from all other IgSF FcR, which see a hinge-proximal site in the C_H2 (or equivalent) domain, the FcαR sees a site bridging the domain boundary between $C_\alpha2$ and $C_\alpha3$, reminiscent of SPA binding (66). Finally, IgA has also been shown to induce eosinophil degranulation via the FcαR, implicating it in antiparasite immunity. Given that many parasites gain access to host tissues by crossing mucosal barriers, this is a logical biologic activity for IgA as well.

IgE

IgE is present in serum in the lowest concentration of all the immunoglobulins. Its rate of synthesis is between 25- and 2,000-fold less than each of the other isotypes, it has the shortest serum half-life, is unable to activate either the classical or alternative complement cascades, and lacks the ability to opsonize antigens. Nonetheless, IgE's biological effects more than compensate for these shortcomings, due to the profound efficiency of its behavior. The principle function of IgE is to arm basophils and mast cells with specific antigen receptors. These cells in turn act as potent dispensers of inflammatory reactions (see Chapter 32).

Plasma cells that produce IgE are chiefly found in the lung and skin. Upon its release from these B cells, circulating IgE is quickly bound by a high-affinity Fcε receptor (FcεRI; $K_D = 10^{-10}$ M) found

on these granulocytes, allowing the IgE molecules to stably remain on the cells for weeks or months. Once primed with many such antigen-specific receptors (recognize that cells bearing FcεRI can have IgE molecules of many different reactivities on their surfaces, unlike the case for antigen-specific B cells), multivalent antigen can then cross-link the bound IgE, indirectly cross-linking the FcεRI molecules as well. Ultimately, this causes mast cells and basophils to release granules containing inflammation-mediating substances and chemoattractants for a variety of cell types. The granule contents of mast cells and basophils are powerful, able to induce rapid responses—including mucous secretion, coughing and sneezing, vomiting, diarrhea, and inflammation. While this type of response can be vital in the clearance of parasites (see Chapter 38), it has the unfortunate consequences of also causing allergy (see Chapter 35) and anaphylaxis in predisposed individuals. In such atopic individuals, it has been seen that increased amounts of IgE are synthesized and found on the surfaces of mast cells and basophils, likely explaining their predilection for these inappropriate responses.

Other cells types also express the high-affinity FcεRI, including Langerhan's cells (171,172) and eosinophils (173), but the rationale for its presence there is yet to be definitively elucidated. In addition, the CD23 surface antigen has also been shown to be a low-affinity IgE receptor (designated FcεRII). Among other cell types, CD23 is known to be expressed on monocytes and some follicular B cells. In fact, monocytes can even be induced to secrete a soluble form of FcεRII (174), but once again the significance of this finding is unclear. Considering the low levels of circulating IgE, the relatively low affinity of the receptor, and the fact that CD23 is known to interact with CD11/CD18, there is doubt as to whether IgE is even an important ligand for this receptor *in vivo*.

Like the pIgR and several of the FcγRs, the FcεRI molecule is also a member of the IgSF (detailed further in the following section). Interaction between IgE and its high-affinity receptor was the first well-characterized Ig-FcR ligand–receptor pair of this type. Originally, studies using synthetic peptides as specific inhibitors of IgE–FcεRI binding identified a 76-amino acid polypeptide spanning $C_ε2$–$C_ε3$ as the FcR recognition site on IgE (175). Subsequently, this localization was refined further to a site in $C_ε3$ analogous to the FcγR site on $C_γ2$. Unlike the binding situation for IgG that was also dependent upon residues in the hinge regions (see above discussion), the extra $C_ε2$ hinge domain of IgE is not believed to play a significant role in the interaction (176).

THE IMMUNOGLOBULIN SUPERFAMILY

Evolution of the Immunoglobulin Superfamily

Soon after the sequencing and structural analyses of antibodies revealed the protein motif of the immunoglobulin domain (7,177), it became apparent that evolution had incorporated Ig homology domains in a variety of other important molecules as well. The sequencing of MHC genes, TCRs, and the pIgR, among others, demonstrated the use of both V region- and C region-type domains by a number of cell-surface proteins of the immune system. Contemporaneously, a number of cell adhesion molecules (CAMs) involved with neurite outgrowth in developing axons were also found to contain Ig-like domains (reviewed in ref. 178). It was quickly recognized that a large family of genes that contained putative immunoglobulin folds existed (2,3,179), whose members were globally implicated in issues of molecular recognition and/or cel-

lular adhesion. Comprised of several multigene families in their own right (V_H, V_L, TCRα, TCRβ, TCRγ, TCRδ, MHC I, MHC II, Sialoadhesin, CAM, etc.), the term *immunoglobulin superfamily* was adopted to refer to this diverse group of genes, which each contained one or more Ig homology domains.

Currently the IgSF encompasses well over 100 genes, and extends across several phylogenetic boundaries (reviewed comprehensively in ref. 4). Disparate species in which IgSF members have been identified include chicken, zebrafish, tunicates, grasshoppers, squid, *C. elegans*, sponges, and *S. cerevisiae*. In addition, reports identifying proteins containing structures reminiscent of immunoglobulin homology domains from prokaryotic organisms (180,181) raise the possibility that this archetypal structure antedates even eukaryote evolution. The discovery of new molecules with novel functional attributes (for this class of proteins) also continues to expand the role of IgSF members. For instance, while the preponderance of immunoglobulin homology domain-containing proteins that have been identified are either cell surface or secreted proteins involved in recognition and adhesion events, a newer class of intracellular muscle proteins (titin, telokin, etc.) belonging to the IgSF demonstrate that the immunoglobulin fold structure is adaptable to an assortment of functional capacities.

The evolution of the IgSF has been the subject of considerable scientific speculation and potentially has implications for both the development of the vertebrate immune system and the process of organogenesis in general. Based upon the overwhelming number of IgSF members that possess adhesive qualities, it has been proposed that the first IgSF molecules were simply single Ig domain extracellular proteins that served as primordial "cellular glues" (2,3). Substantiating this argument is the noted stability of the compact β-barrel structure of the immunoglobulin fold, which would foster its utility in harsh extracellular environments. Further bolstering this hypothesis is the fact that numerous IgSF proteins participate in both homotypic and heterotypic interactions with other IgSF molecules, demonstrating their potential to act as adhesion molecules.

Some evidence suggests that IgSF proteins have promoted clustering of cells since the earliest stages of eukaryotic development. For example, the yeast *S. cerevisiae* uses the IgSF glycoprotein α-agglutinin to mediate cell–cell contact during mating (182). IgSF forebears may have also allowed the first examples of rudimentary organogenesis in phylogeny. The slime mold *Dictyostelium*, which bridges the gap between unicellular and multicellular eukaryotes, uses a protein possessing a region with striking similarity to an immunoglobulin domain for the purpose of forming aggregations called "fruiting bodies" when conditions are nutrient-scarce (183,184).

Finally, it has also been put forth that ancestral IgSF glycoproteins may have mediated the first evolutionary examples of allorecognition in colonial invertebrates (185). In defense of this proposition, it is noteworthy that two examples of metazoan receptor tyrosine kinases with purported recognition functions have been shown to possess extracellular Ig-like domains, one from the cnidarian *Hydra vulgaris* and the other from the marine sponge *Geodia cydonium*. From these data, then, it is possible to make tentative, yet tenable, conjecture that the complex cellular and molecular interactions of the vertebrate immune system (mediated in no small part by members of the IgSF) may in fact be an outgrowth of this primitive allorecognition. In this light, the notable analogy between vertebrate graft acceptance/rejection reactions and colonial invertebrate fusion/rejection phenomena perhaps takes on new significance (see Chapter 18).

Regardless of its derivation, the immunoglobulin domain has obviously proven to be a pliable evolutionary substrate, amenable to mutation and diversification for a number of important reasons. First, as was noted for the actual domains of immunoglobulins, the primary structure of these units can vary dramatically without appreciably altering their tertiary structure (186,187). This is particularly evident in the interconnecting loops that join the β strands, allowing them to diverge rapidly to perform a multitude of distinct functions. Second, most Ig domains are encoded by discrete exons, facilitating their duplication by relatively simple genetic events. This one-domain-per-exon rule is also conducive to alternative splicing phenomena, encouraging differential expression of IgSF molecules as well. This is further accommodated by a splicing convention followed by most IgSF exons: The 3' end of one exon is always the first position of a codon, while the 5' end of the next tandem unit begins with the second position of a codon. Thus, immunoglobulin homology domains of IgSF proteins may be easily duplicated in tandem (the C. elegans muscle protein twitchin contains 26 Ig-like domains) and shuffled to create new genes with the capacity to diversify both somatically and evolutionarily. Finally, the propensity of Ig domains to form homotypic and/or heterotypic dimers (also demonstrated by immunoglobulin proper) forms the basis for proteins which serve as receptor and ligand molecules. These combinatorial associations enhance their diversification potentials still further.

Despite the inherent complexity in a gene superfamily containing such vast disparities in its members' functional qualities, it was recognized early on that IgSF proteins could be subdivided into distinct "sets" on the basis of sequence and structural analyses (2). These groupings are based upon the arrangement of the β strands of the immunoglobulin fold and are schematized in Fig. 21. Note

that while V-set domains are composed of a four-strand sheet and a five-strand sheet (as detailed earlier), C-set domains have four-strand and three-strand layers; these are discriminated on the basis of placement of the D strand in the sheet of strands A, B, and E (the C1 set) or with the layer formed by strands G, F, and C (the C2 set). However, studies with the IgSF muscle protein telokin have revealed a new "I set," which has domain features that are intermediary between the V and C1 sets (188). Moreover, many IgSF adhesion molecules and cell-surface receptors likely belong to this I set, rather than to the sets to which they were previously ascribed. In any case, the IgSF remains a fascinating collection of proteins with structural similarities but a wide array of functional abilities. While immunoglobulin remains the definitive example of this class of molecules, a number of IgSF proteins are also of particular significance to humoral immune responses, and their structural and functional characteristics are briefly summarized in the following sections.

Fc Receptor Molecules

FcR allow antibodies to interact with cells of both the specific and non-specific immune systems. In so doing, FcR connect humoral immune responses to cellular immune responses, and more globally, acquired immunity to that of innate immunity. These contacts play two vital roles in the biology of immune functioning. First, FcR allow antibodies to act as "flags" signaling the need for certain cellular effector events, such as phagocytosis and ADCC. Second, the different FcR facilitate antibody acting as a mediator of overall immune regulation. The signals they transmit can induce changes in cytokine secretion, expression of cell-surface receptors, and extensive differentiation programs (189).
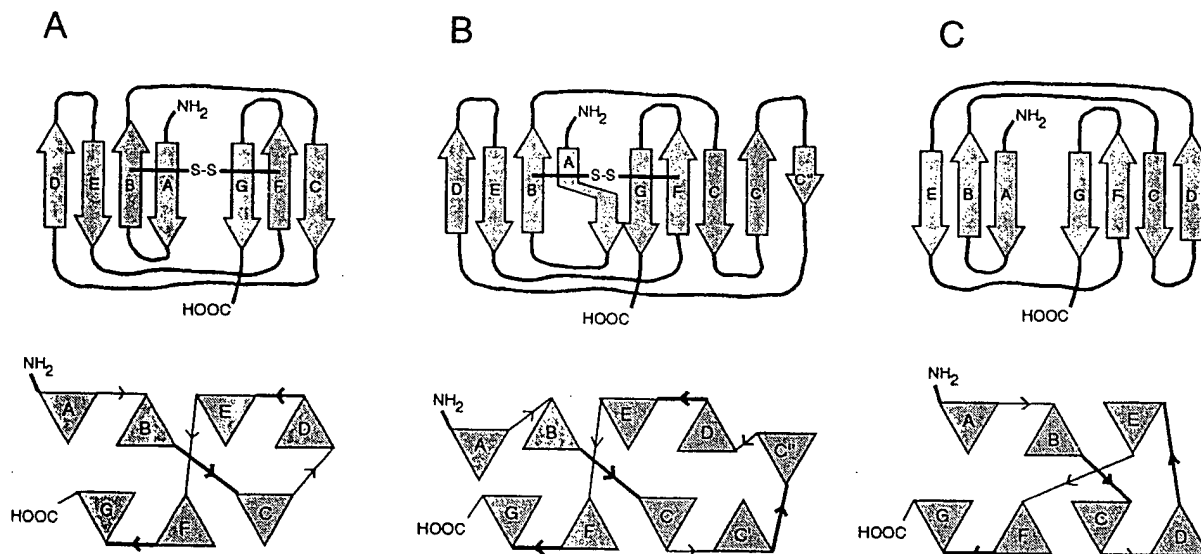
FIG. 21. Topology of different immunoglobulin domain types. Diagrams of the (A) C1 set, (B) V set, and (C) C2 set are presented. In the upper part of the figure, β strands are depicted as *broad arrows* and their intervening loops by *thin lines*. Note that the V-type domains have five- and four-stranded faces, while C1- and C2-type domains have four- and three-strand faces. The C region-like structures are discriminated on the basis of placement of their D strand. In the lower part of the figure, an end-on view of the different β-barrels is shown. *Triangles (with their apex at the top)* symbolize β strands running out of the plane of the paper; *triangles (whose apex points down)* are β strands traveling into the paper. *Bold lines* represent connecting loops at the top of the immunoglobulin fold; *thin lines* indicate connections at the bottom of the domain. (From ref. 4, with permission.)

Three large classes of molecules can bind Fc regions: glycosyltransferases, which recognize oligosaccharide derivatives on antibodies, lectin-like molecules, and receptors belonging to the IgSF. Of the "true" FcR that recognize antibody protein determinants rather than carbohydrate, all FcR thus far identified belong to the IgSF, other than the low-affinity IgE receptor (CD23/FcεRII). These molecules include FcγR I, II, and III (CD64, CD32, and CD16), FcεRI, FcαR (CD89), and the pIgR, which has already been discussed. All cells of lymphoid origin express FcRs, although the profiles and isotype specificities between lineages can vary greatly (reviewed in refs. 190 and 191). While receptors for all classes of immunoglobulin have been described as biological activities, human FcμR and FcδR have not yet been cloned. Thus far, the FcγR proteins and FcεRI are the most well characterized examples of these molecules.

FcγR Molecules

Receptors for the Fc portion of IgG are of three types (reviewed in ref. 192). FcγRI (CD64) is a high-affinity receptor and the only one able to bind monomeric IgG. It possesses three extracellular Ig-like domains. FcγRII (CD32) and FcγRIII (CD16) are both low-affinity receptors that bind IgG-containing immune complexes. They each have only two extracellular Ig homology domains. Schematic diagrams of the FcγRI, FcγRII, FcγRIII complexes, along with the FcεRI and FcαR complexes, are presented in Fig. 22.

FcγRI is a 70-kD glycoprotein that is constitutively expressed at low levels on monocytes and macrophages. IFN-γ upregulates its levels on these cells, and also can induce its expression by neutrophils. FcγRI's affinity for IgG is highest for the IgG1 and IgG3 subclasses ($K_D = 10^{-8}$ M), tenfold lower for IgG4, and will not bind IgG2. Functionally, the primary effect of cross-linking FcγRI molecules appears to be the potentiation of both ADCC and phagocytosis. As IFN-γ enhances both of these activities by the cell types known to express FcγRI, this would fit well with their being important roles for the receptor.

Like surface immunoglobulin, FcγRI requires accessory proteins in order to transmit signals. This is, in fact, a common feature of most FcRs (except for FcγRII) and an interesting parallel between the IgSF antigen receptors (BCR and TCR) and the IgSF "indirect"

antigen receptors (the FcR), which use antibody to bridge the span between FcR and antigen (193). In the specific case of FcγRI, the actual signaling molecule is a 12γ-kD transmembrane protein designated the "γ-subunit" or, more generally, FcRγ. This nomenclature can be particularly confusing, as the "γ" of FcRγ refers not to the fact that it is part of the of FcγR complex (the receptor for γ-class immunoglobulin), but rather to γ as an individual subunit of a multi-molecule complex. In any case, for the FcγRI *complex*, the α subunit is the actual IgSF protein FcγRI, and the γ subunit is FcRγ, which forms a disulfide-linked homodimer. Complicating terminology further, FcRγ is also a subunit of other FcR complexes, including that of the FcγRIIIA and those of the non-FcγR, FcεRI and FcαR. Intriguingly, FcRγ is a close homologue of the TCR-associated protein CD3ζ. In fact, CD3ζ cannot only heterodimerize with FcRγ, but also has been shown to be capable of functionally substituting for FcRγ as the signal-transducing subunit of the FcγRIIIA receptor complex (194).

The situation for FcγRII (CD32) is even more complex. FcγRII is the product of three distinct but homologous genes: FcγRIIA, FcγRIIB, and FcγRIIC. This is further complicated by the fact that at least two of the FcγRII genes are alternatively spliced to generate multiple isoforms (195). The FcγRIIA gene gives rise to two transcripts: FcγRIIa1, which has a transmembrane domain, and FcγRIIa2, which lacks it. The FcγRIIB gene has three isoforms— FcγRIIb1, FcγRIIb2, and FcγRIIb3—generated by differential splicing and alternative polyadenylation processing. Collectively, the FcγRII variants are the most ubiquitously expressed FcγRs, being present on monocytes, macrophages, neutrophils, B lymphocytes, megakaryocytes, and platelets. Specifically, megakaryocytes express FcγRIIA (both isoforms), B lymphocytes express FcγRIIB (b1 and b2 transcripts) and FcγRIIC, and cells of myelomonocyte derivation produce at least one or more isoforms from all three genes (195).

Functionally, due to their expression on many cell types, FcγRII signals cause diverse effects. When cell surface FcγRII engage IgG immune complexes (all subclasses, with varying affinities), they potentiate several biologic changes, most immunoregulatory in nature. Generally, these signals down modulate IgG-, IgA-, and IgE-mediated activations of a number of cell types, including monocytes and macrophages, granulocytes, mast cells, and Langerhans and other dendritic cells. They also induce platelet aggregation at the site of immune complexes and effect B cell feedback inhibition by down-
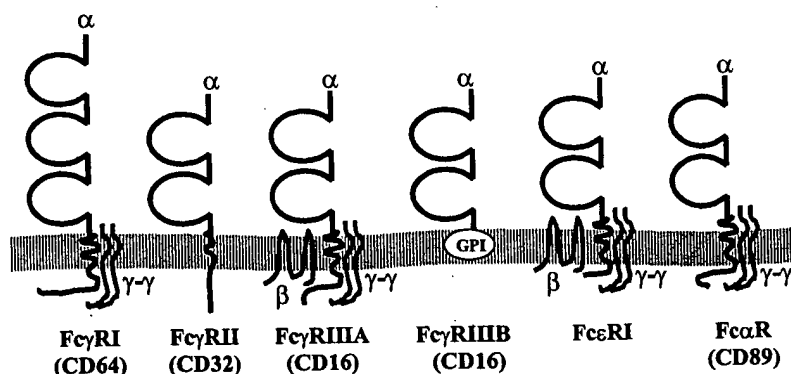
FIG. 22. Schematic diagram of human immunoglobulin Fc receptors belonging to the IgSF. Each Ig domain is depicted as a *rounded bulge*. The α chains are the components of the receptor complex that determine binding specificity. β and γ chains are responsible for association and signal propagation by the receptor(s).

regulating both proliferation and antibody production. Specifically, FcγRIIB are known to suppress BCR-mediated activation signals when the two are coaggregated (196). The signal for this inhibition is brought about in a manner unique among all IgSF FcR. As single-chain receptors without accessory proteins, FcγRII are able to transduce their own signals. They do so by way of an immunoreceptor tyrosine-based inhibition motif (ITIM) present in the cytoplasmic region of the protein (197).

Specific binding of IgG by FcγRII was initially mapped to the C-terminal portion of the second extracellular Ig-like domain of the receptor (198). This primary site, comprising residues Asn 154–Ser 161, has since been revised to also include domain 2 stretches Ser 109–Val 116 and Phe 129–Thr 135, along with domain 1 contacts (199). In sum, a three-dimensional model of the entire FcγRII extracellular region predicts that loops of both Ig-like domains that co-localize to the domain interface are responsible for the recognition of IgG.

The final IgG FcR is the FcγRIII (CD16). FcγRIII has two extra-cellular Ig-like domains and is encoded by two separate genes whose expression varies by cell type. On monocytes, macrophages, and NK cells, it is a transmembrane glycoprotein called FcγRIIIA. The FcγRIIIA receptor protein has three accessory proteins with which it is complexed. The first is a 30-kD "β-subunit" having four transmembrane regions, which is also a component of the FcεRI complex. The other protein(s) associated with FcγRIIIA is a homodimer of FcRγ subunits, which, as explained earlier, is also part of the receptor complexes of FcγRI, FcεRI, and FcαR (200). The other FcγRIII gene encodes a glycophosphoinositol-linked protein termed FcγRIIIB that is expressed on neutrophils. Individuals deficient in this gene suffer from a condition called paroxysmal nocturnal hemoglobinuria, characterized by increased susceptibility to infection and delayed clearance of immune complexes (201). FcγRIII's binding preference is for IgG1 and IgG3, both of which it binds with low affinity only in the form of immune complexes. Biologic activities fostered by FcγRIII include ADCC, phagocytosis, and transport of internalized Ab-Ag complexes to the antigen-presentation pathway.

FcεRI and FcαR Molecules

The high-affinity IgE receptor (FcεRI) was the first and best characterized of the FcR (202). FcεRI is a transmembrane protein having two extracellular Ig homology domains and, like previous examples, associates with accessory proteins for signal-transmission purposes. It is expressed on mast cells, basophils, eosinophils, Langerhans cells, and on the monocytes of atopic individuals (203). The proteins associated with FcεRI are the same as for FcγRIIIA: one β subunit and a homodimer of FcRγ proteins. Gene targeting experiments to verify the roles these proteins play in the FcεRI complex have yielded reassuring results: homozygous deletion of the IgSF chain of the receptor created animals that were predictably resistant to anaphylaxis (204); disruption of FcRγ caused the same phenotype, plus defects in ADCC and phagocytosis consistent with the γ-subunit's participation in other receptor complexes (205,206).

The Fcε-binding site on FcεRI has been localized to three regions of the second extracellular Ig domain (198). It is important to remember that serum IgE binds to the receptor in a high-affinity interaction not dependent on antigen (unlike all other antibody isotypes, which must first bind antigen in order to be recognized by their respective FcR). This allows polyvalent immunogens to bind effector cells directly, without the need for conformational change of the immunoglobulin and/or immune complex formation. This contributes to the rapidity of the response exhibited by cells expressing FcεRI. The specific biologic effects of cross-linking the receptor were discussed previously in the section on IgE function.

The final FcR of the IgSF to be covered is that for IgA. The FcαR (CD89) is the most recently identified and least characterized of the different Ig receptor classes. It possesses two extracellular Ig domains and is expressed by monocytes, macrophages, neutrophils, and eosinophils. Several isoforms have been identified: a cell surface FcαRa form, which has intracellular and transmembrane domains, an FcαRb form lacking these domains that is both secreted and associated with the cell surface (by an unknown mechanism), and even an isoform lacking the membrane-proximal Ig-like domain. Structurally, the FcαR has homology with the FcγR molecules. Like FcγRI, FcγRIII, and FcεRI, the IgA receptor complex includes the FcRγ homodimer as a signaling component. The particulars of the receptor binding site on the FcαR are not yet determined. The site it recognizes on IgA and the effects mediated by FcαR-binding were detailed in the section on IgA function.

Coreceptor CD4 and CD8 Molecules

Antigen-recognition functions in the body are not limited to immunoglobulin but are also performed by receptors on T cells (TCR), which bind antigenic peptides. In a defining event of immune responses, these antigenic fragments are "presented" to T cells within the context of molecules of the MHC. Recognition of MHC/antigen by TCR is the fundamental biologic interaction responsible for initiation, perpetuation, and mediation of cellular immunity. As pertains to antibody, binding of MHC/peptide complexes by TCR is also vital for recruitment of T cell help needed in many humoral immune responses. While the details and effects of these vital interactions are well beyond the scope of this chapter (see Chapters 8–13), many IgSF proteins play key roles in assuring its productive functional outcome. Of these, the TCR "coreceptors" CD4 and CD8 are crucial components worthy of mention here.

CD4 and CD8 were among the first non-immunoglobulin IgSF members for which structural information became available (reviewed in ref. 207). These molecules are each expressed on the surface of T cells, where they participate in TCR/MHC interactions (schematized in Fig. 23) by engaging non-polymorphic regions of the MHC in low-affinity interactions (208,209). T cells break down into two major functional subclasses—helper T cells and cytotoxic T cells—characterized by different responses to antigen. Both T cell types utilize the same group of TCR genetic elements to compose their specific antigen-receptors, however. Ordinarily, although there are notable exceptions, T cell effector functions correlate with the type of MHC protein with which they interact. MHC I molecules specify cytotoxic T cell responses and are found on most cell types of the body. MHC II molecules, on the other hand, dictate helper T cell functioning and are more restricted in their expression, typically found only on "professional" antigen-presenting cells. CD8 molecules bind to class I MHC proteins, while CD4 molecules mediate interaction with class II; thus, these two proteins play an important role in determining what type of response a particular T cell is likely to mediate.
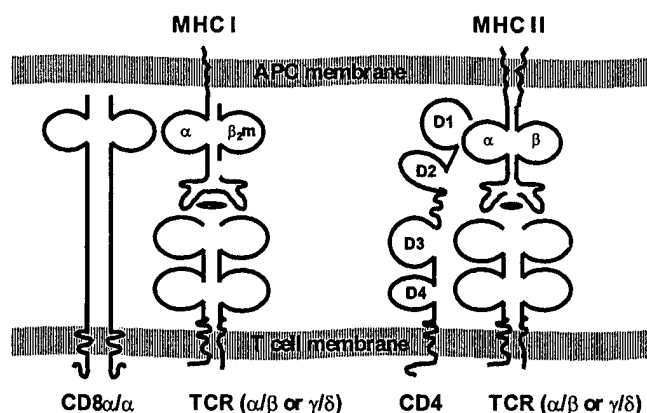
FIG. 23. Schematic diagram of the CD8 and CD4 coreceptor molecules. The figure shows both TCR/MHC coreceptor complexes on the same membrane, although T cells express either CD8 or CD4 for the majority of their lifetime. *Bulges* represent Ig domains, and *gray ovals* signify peptide presented by MHC molecules. Only the CD8α/α homodimer is schematized here, although the CD8α/β heterodimer presumably binds MHC I in similar fashion. The models demonstrate the simplest stoichiometry for association; other possible modes of interaction are discussed in the text.

Both CD8 and CD4 are glycoproteins and, in their most common incarnations, possess transmembrane segments and short cytoplasmic tails (210). The intracytoplasmic regions of both molecules interact with the src-like tyrosine kinase p56^{lck} (211), which presumably serves to allow signal-transduction necessary for proper thymic selection (212) and activation (213) of T cells. Reinforcing this idea is the fact that co-ligation of CD4 or CD8 with TCR greatly enhances stimulation of T cells relative to that of cross-linking TCR alone (see Chapters 12 and 13). In addition, CD8 and CD4 also increase the avidity of interaction between TCR and MHC, by virtue of their action as adhesion molecules between the two cell membranes. Collectively, these behaviors have been estimated to boost antigen recognition over 100-fold from that of basal levels (TCR engagement by MHC/Ag only). However, despite these similarities in their biologic effects, structurally CD8 and CD4 have many important differences. While many details are still unresolved, a number of crystals involving these two proteins have been solved by x-ray diffraction, permitting a thorough examination of their salient characteristics.

CD8

CD8 exists as a disulfide-linked dimer in one of two forms. A homodimer of two CD8α subunits was the first human isoform identified (214). Subsequently, a CD8α/β heterodimer was described as well (215,216). Both proteins are 34-kD and have homologous (although only 17% identical) N-terminal Ig-like domains, extended hinge regions of 50 (α) and 30 (β) amino acid residues, single transmembrane stretches, and short cytoplasmic tails. While the CD8β chain lacks residues necessary for interaction with p56^{lck}, the heterodimer is still capable of interacting with it via the α subunit. Nevertheless, several lines of evidence indicate that a specific role for the heterodimer may exist, apart from that of the homodimer. Differences in functional effects (217), thymic selection (218), avidity for MHC (219), and p56^{lck} activity (220) have all been attributed to the

CD8β chain. In addition, there are differences in expression of the different isoforms: thymocytes and peripheral T cells are CD8α/β+, TCR γ/δ+ intraepithelial lymphocytes (IEL) are primarily CD8α/α+, and TCR α/β+ IEL express either of the two molecules. Finally, investigations into CD8β have divulged that there are actually two genes encoding this protein (221) (i.e., the locus has been recently duplicated), which—along with alternative splicing phenomena—results in as many as seven unique CD8β isoforms being expressed (222). Four of these transcripts lack the transmembrane region of the message, raising the possibility that some forms of CD8β may be secreted. In sum, the story of this protein subunit, and of the CD8 heterodimer that derives from it, is still an active area of research that is yet to be clarified.

The enumeration of the CD8α/α homodimer, thanks in large part to two crystal structures, is somewhat further along. The primary advance in this regard was the solving of the amino-terminal domains of human CD8α/α (223). This study revealed that these 113-amino acid segments formed V-type Ig domains consisting of four- and five-strand layers (see Fig. 24). In agreement with its V region-type topology, CD8α domains were shown to dimerize with one another via their five-strand faces, as do immunoglobulin V domains (see Colorplate 7). Two significant structural disparities between CD8α and immunoglobulin V regions were also recognized. First, the C'-C'' loop (corresponding to CDR2) is extended in the α subunit (note the right side of Fig. 24D). Second, while the usual intradomain disulfide bridge between β strands B and F was identified, an unpaired cysteine in strand C is also conserved. In rodent CD8α this residue has been shown to form the intradomain cystine together with the Cys of strand B (224). The α subunit hinge region is extensively glycosylated (via O-linkages), and this is thought to promote its adopting an extended structure (223,225–227). This is particularly important because an elongated conformation of the hinge would be necessary to allow the N-terminal Ig domain to interact with the MHC I molecule appropriately (refer back to Fig. 23).

The other important structural features of CD8 concern its interaction with class I MHC molecules. Initial experiments into these questions indicated that the CDR-like loops of CD8α were involved with recognizing a negatively-charged region on the α3 domain of MHC I (228). The aforementioned crystal structure supported this conclusion by demonstrating that these same loops were the only region on CD8 where positive charge was localized. Mutational studies performed after the crystal was solved also implicated residues in the A and B strands of the CD8α protein as contact points with the α2 domain of MHC I (229). As each CD8 homodimer has two α chains, and as the A and B strands of each chain are not on the dimerizing face of the subunit, this implied that CD8's interaction with MHC I could be bivalent (i.e., one CD8α/α and two MHC class I proteins). The publication of a crystal structure of the complex containing CD8α/α and MHC I plus peptide has seemingly resolved these issues (230). The homodimer was shown to have contacts with not only the α3 domain, but also the α2 domain, and even with the β2-microglobulin subunit of MHC class I. Strikingly, the negatively-charged region of the α3 MHC I protein fits between the CDR-like loops of the two CD8α subunits in the fashion of classical antibody–antigen interactions! However, because both CD8 subunits are needed for the binding of one such loop, it would appear that the stoichiometry of the CD8/MHC I interaction is in fact 1:1. Because the clustering of receptor complexes is likely an important feature for the generation of intracellular signals, this is a crucial piece of information, as shall be seen for CD4.
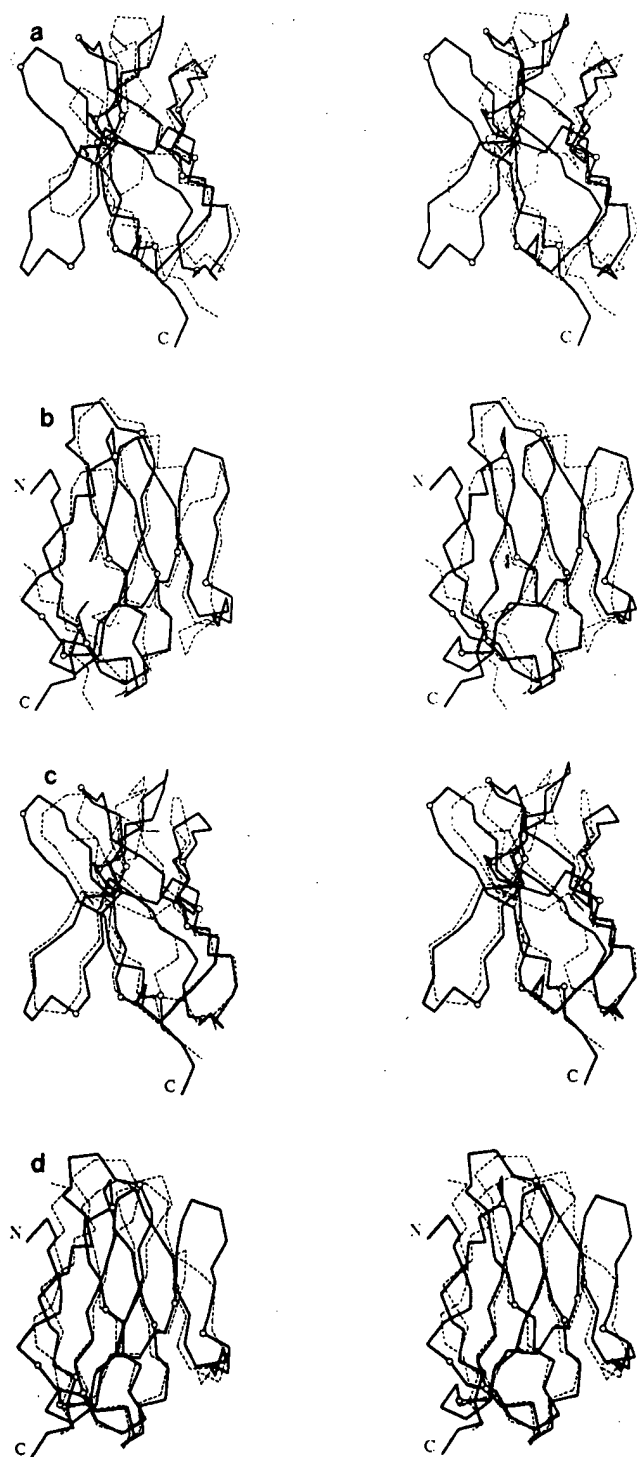
FIG. 24. Stereoviews of the α-carbon backbones of CD8α, CD4 domain 1, and the V_L of the antibody REI. In **(a,b)** the Ig domain of CD8α (*solid lines*) is superimposed on domain 1 of CD4 (*dashed lines*). In **(c,d)** V_L (now in *dashed lines*) is overlaid by CD8α. Parts **(a)** and **(c)** are side views (parallel to the dimerization surface); **(b)** and **(d)** are perpendicular to the β-sheet faces. In all cases, the CDR loops are at the top of the figure. Comparing parts **(a)** to **(c)** illustrates the truncation of the F–G (*top left*) and C–C′ (*bottom left*) loops of CD4 D1 relative to CD8α and V_L. Comparison of the N-termini [left edge of **(b)** and **(d)**] shows the shortening of CD4 D1's A strand as well. The CDR2-like C′-C″ loop [upper right of **(b)** and **(d)**] demonstrates that this segment is elongated in both CD8α and CD4 D1 relative to V_L. (From ref. 207, with permission.)

CD4

CD4 has four extracellular Ig homology domains (D1–D4) and is thought to exist as a 55-kD monomer on the cell surface (231). Scientifically, CD4 became the center of intense scrutiny when it was demonstrated to be the molecule utilized by HIV for attachment to T cells. Like immunoglobulin, proteolytic analyses established that CD4 generated stable fragments upon cleavage. These, in turn, were the initial substrates for crystallographic study. The amino-terminal (and T cell membrane-distal) D1D2 segments were the first regions of CD4 structurally determined, as they had been shown to contain the HIV-binding site. These studies (232,233) described an N-terminal V-type Ig domain (D1) and a smaller, unusual Ig-like domain (D2), each with features unique among previously reported IgSF structures.

D1 is a four- and five-strand domain (see Colorplate 7) that maintains the normal core intradomain disulfide bond and its associated residues. However, by comparison with immunoglobulin V regions, it became apparent that part of the A strand was missing (see Fig. 24). Similarly, the two loops connecting β strands C to C′ and F to G were both shortened in length. Since amino acids in these positions of immunoglobulin (and CD8) participate in dimerization events, this was taken to be reflective of the fact that CD4 was not known to dimerize. Like CD8, the CDR2-homologous C′-C″ loop of CD4 D1 is also extended relative to immunoglobulin (compare the right edges of Figs. 24B and 24D). In fact, a Phe residue found on this lengthened segment has been shown to be crucial for binding of HIV gp120 to CD4. Another interesting characteristic of these crystals concerns the D1 to D2 domain connection. Note in Colorplate 7 that the G strand of D1 is contiguous with D2's A strand (contrast with the elbow peptides connecting V and C domains). As a result, D1 and D2 have a large amount of longitudinal contact and, accordingly, little flexibility to move relative to one another.

The D2 domain is even more peculiar. D2 is smaller than most Ig domains, and it consists of only seven β strands, like Ig C-set domains (see Colorplate 7). Unlike the canonical C-type domain, D2 has switched the placement of one of its strands from one face to the other (i.e., it belongs to the C2 set; refer back to Fig. 21). Remarkably, D2 also fails to conserve the core residues necessary for the typical intradomain cystine. Curiously, its intradomain disulfide linkage is between cysteines found in the same β-stranded sheet (i.e., its disulfide bond is *intra*sheet instead of the usual *inter*sheet). Despite this idiosyncratic arrangement, D2 forms an Ig fold consistent with other Ig domains—a powerful testimony to the principle of tertiary structural conservation in the face of primary structural variation, embodied by the IgSF.

The remaining two Ig domains of CD4—D3 and D4—have also been crystallized (albeit the rat CD4 homologue) and their structures elucidated by x-ray diffraction (234,235). Remarkably, the D3D4 fragment adopts a conformation resembling that of the D1D2 portion of the molecule (see Colorplate 7), embracing prior hypotheses that CD4 arose by way of duplication of a two-domain precursor (236). D3 is a larger V-type domain homologous to D1, and D4 is a smaller module reminiscent of D2. Once again, the D3 G strand is contiguous with the A strand of D4, closely approximating the two domains and limiting the flexibility between them. However, D3 does have characteristics which distinguish it from the D1 domain. First, D3 fails to conserve the intradomain disulfide bond, resulting in a "relaxed" domain, which is packed less-tightly. Second, unlike D1, the C to C′ and F to G loops of D3 are

not shortened relative to immunoglobulin V regions. Still, these loops are unlikely to mediate dimerization of D3 domains (as they did in the Fv fragment of immunoglobulin), on account of an N-linked glycosylation site on the F strand of D3's inner face that would interfere with such interactions. Recently, a recombinant soluble form of human CD4 has also been solved crystallographically (237). While consistent with previous conclusions, this report makes the added contribution of definitively establishing the D1D2 to D3D4 junction as a hinge-like region of the protein. Thus, the rod-like two domain portions of CD4 (D1D2 and D3D4) are able to bend at a point of flexion akin to the scenario for Fab–Fc bending at the antibody hinge.

Studies have also examined the means by which CD4 binds to MHC II proteins (refer back to Fig. 23). CD4 contacts the α2 and β2 domains of class II molecules using a variety of residues in the D1 and D2 domains. Most evidence points to a large surface of CD4, involving both lateral faces of D1 and the F–G loop of the D2 domain, being implicated in MHC II binding (238–240). Once more the question of the stoichiometry of interaction between coreceptor and MHC is of interest. As both sides of the D1 domain appear to contact the class II molecule, the prospect of a bivalent complex (one CD4 protein with two MHC II molecules) is once again at issue. Given that crystal studies of class II proteins have demonstrated a dimeric association between MHC II molecules, this seems a plausible mode of complex formation. Contrarily, the crystal of soluble human D1–D4 (237) has revealed a homodimeric association between D4 domains of CD4. This implies that—as had been proposed by others (241)—opposite faces of a CD4 dimer may interact with two separate class II molecules. Regardless of the specifics of their dimeric interactions, it is reasonable to conclude that multiple surfaces of CD4 are responsible for binding, and that the majority of these amino acids reside in the D1 domain.

CONCLUSION

Although these discussions only scratch the surface of structure–function relationships within the immunoglobulin superfamily of proteins, it is hoped that this chapter has served to introduce the reader to the inherent utility—and exquisite beauty—of the immunoglobulin domain as both an evolutionary tool and molecular motif. While antibody proteins have been structurally characterized and functionally probed to an unparalleled degree by the concerted and persistent efforts of the scientific community, the continued emergence of unexpected findings indicates that a great wealth of knowledge is yet to be tapped in their inquiry. Moreover, the TCR and MHC IgSF proteins occupy an even greater role in terms of immune system functioning, and their investigations have been fruitful fields for study, as well. Other molecules, like CD4 and CD8, while not the focus of attention for the prolonged duration that has been the case for immunoglobulin, have nonetheless seen seminal findings in the pursuit of their understanding, and have served to broaden our comprehension of the IgSF's variety and capacity. Still others, like the antigen-receptor signaling proteins Igα, Igβ, and CD3 subunits, or the co-stimulatory molecules B7-1 (CD80), B7-2 (CD86), CTLA-4, and CD28 have only recently been described in detail and will no doubt be centers of concentration in the immediate future. Given the unequivocal fact that new and exciting IgSF members are yet to be discovered, when one considers the pervasiveness of this class of proteins in

immunology—and in biology as a whole—it is perhaps accurate to surmise that the study of the immunoglobulin superfamily is still only in its infancy.

REFERENCES

1. Dreyer WJ, Bennett JC. The molecular basis of antibody formation: a paradox. *Proc Natl Acad Sci USA* 1965;54:864–869.
2. Williams AF, Barclay AN. The immunoglobulin superfamily—domains for cell surface recognition. *Annu Rev Immunol* 1988;6:381–405.
3. Hunkapiller T, Hood L. Diversity of the immunoglobulin superfamily. *Adv Immunol* 1989;44:1–63.
4. Brummendorf T, Rathjen FG. Cell adhesion molecules 1: immunoglobulin superfamily. *Protein Profile* 1995;2:963–1108.
5. Hilschmann N, Craig LC. Amino acid sequence studies with Bence-Jones proteins. *Proc Natl Acad Sci USA* 1965;53:1403–1409.
6. Silverton EW, Navia MA, Davis DR. Three-dimensional structure of an intact human immunoglobulin. *Proc Natl Acad Sci USA* 1977;74:5140–5144.
6a. Carayannopoulous L, Capra JD. Immunoglobulins: structure and function. In: Paul WE, ed. *Fundamental Immunology*, 3rd ed. New York: Raven Press Ltd., 1993.
7. Hill RL, Delaney R, Fellows RE, Lebovitz HE. The evolutionary origins of the immunoglobulins. *Proc Natl Acad Sci USA* 1966;56:1762–1769.
8. Edmundson AB, Ely KR, Abola EE, Schiffer M, Panagiotopoulos N. Rotational allomerism and divergent evolution of domains in Ig light chains. *Biochemistry* 1975;14:3953–3961.
9. Davies DR, Padlan EA, Segal DM. Immunoglobulin structures at high resolution. *Annu Rev Biochem* 1975;44:639–667.
10. Amzel LM, Poljak RJ. Three dimensional structure of immunoglobulins. *Annu Rev Biochem* 1979;48:961–997.
11. Deisenhofer J, Colman PM, Epp O, Huber R. Crystallographic structural studies of a human Fc fragment. *Hoppe-Zeylers Z Physiol Chem* 1976;357:1421–1434.
12. Deisenhofer J. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9 and 2.8 Å resolution. *Biochemistry* 1981;20:2361–2370.
13. Federation of American Society for Experimental Biology Symposium, "Two Genes, One Polypeptide Chain." *Fed Proc* 1972;31:176–209.
14. Porter RR. Separation and isolation of fractions of rabbit γ-globulin containing the antibody and antigenic combining sites. *Nature* 1958;182:670–671.
15. Nisonoff A, Wissler FC, Lipman LN. Properties of the major component of a peptic digest of rabbit antibody. *Science* 1960;132:1770–1771.
16. Palmer JL, Nisonoff A. Dissociation of rabbit γ-globulin into half-molecules after reduction of one labile disulfide bond. *Biochemistry* 1964;3:863–869.
17. Fleischman JB, Porter RR, Press E. The arrangement of the peptide chains in γ-globulin. *Biochem J* 1963;88:220–228.
18. Nisonoff A, Thorbecke GJ. Immunochemistry. *Annu Rev Biochem* 1964;33:355–402.
19. Natvig JB, Kunkel HG. Human Igs: classes, subclasses, genetic variants, and idiotypes. *Adv Immunol* 1973;16:1–59.
20. Nisonoff A, Hopper JE, Spring SB. *The antibody molecule.* New York: Academic Press, 1975.
21. Kindt TJ, Capra JD. *The antibody enigma.* New York: Plenum Press, 1984.
22. Hasemann CA, Capra JD. Immunoglobulins: structure and function. In: Paul WE, ed. *Fundamental immunology*, 2nd ed. New York: Raven Press, 1989.
23. Weir DM, ed. *Handbook of experimental immunology*, vol. 3, *Genetics & molecular immunology.* London: Blackwell Scientific Press, 1986.
24. Williams RC Jr, Kunkel HG, Capra JD. Antigenic specificities related to the cold agglutinin activity of gamma M globulins. *Science* 1968;161:379–381.
24a. Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C. *Sequences of proteins of immunological interest.* Washington, DC: US Department of Health and Human Services, 1991.
25. Wu TT, Kabat EA. An analysis of the sequences of the variable regions of Bence-Jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med* 1970;132:211–250.
26. Capra JD, Kehoe JM. Hypervariable regions, idiotypy, and the antibody-combining site. *Adv Immunol* 1975;20:1–40.
27. Colman PM. Structure of antibody-antigen complexes: implications for immune recognition. *Adv Immunol* 1988;43:99–132.
28. Chothia C, Novotny J, Bruccoleri R, Karplus M. Domain association in immunoglobulin molecules. *J Mol Biol* 1985;186:651–663.
28a. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 1987;196:901–917.
29. Tonegawa S. Somatic generation of antibody diversity. *Nature* 1983;302:575–581.
30. Meier JT, Lewis SM. P nucleotides in V(D)J recombination: a fine-structure analysis. *Mol Cell Biol* 1993;13(2):1078–1092.
31. Meek KD, Hasemann CA, Capra JD. Novel rearrangements at the immunoglobulin D locus. *J Exp Med* 1989;170:39–57.
32. McKean D, Huppi K, Bell M, Staudt L, Gerhard W, Weigert M. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc Natl Acad Sci USA* 1984;81:3180–3184.

33. Neuberger MS, Milstein C. Somatic hypermutation. *Curr Opin Immunol.* 1995; 7:248–254.

34. Croce CM, Shander M, Martinis J, et al. Chromosomal location of the genes for human immunoglobulin heavy chains. *Proc Natl Acad Sci USA* 1979;76: 3416–3419.

35. Malcolm S, Barton P, Murphy C, Ferguson-Smith MA, Bentley DL, Rabbitts TH. Localization of human immunoglobulin kappa light chain variable region genes to the short arm of chromosome 2 by *in situ* hybridization. *Proc Natl Acad Sci USA* 1982;79(16):4957–4961.

36. Erickson J, Martinis J, Croce CM. Assignment of the genes for human λ immunoglobulin chains to chromosome 22. *Nature* 1981;294:173–175.

37. D'Eustachio P, Pravtcheva D, Marcu K, Ruddle FH. Chromosomal location of the structural gene cluster encoding murine immunoglobulin heavy chains. *J Exp Med* 1980;151(6):1545–1550.

38. Swan D, D'Eustachio P, Leinwand L, Seidman J, Keithley D, Ruddle FH. Chromosomal assignment of the mouse κ light chain genes. *Proc Natl Acad Sci USA* 1979;76(6):2735–2739.

39. D'Eustachio P, Bothwell AL, Takaro TK, Baltimore D, Ruddle FH. Chromosomal location of structural genes encoding murine Ig λ light chains. *J Exp Med* 1981;153(4):793–800.

40. Schroeder HW Jr, Hillson JL, Perlmutter RM. Structure and evolution of mammalian V_H families. *Int Immunol* 1990;2:41–50.

41. Kirkham PM, Mortari F, Newton JA, Schroeder HW Jr. Immunoglobulin V_H clan and family identity predicts variable domain structure and may influence antigen binding. *EMBO J* 1992;11:603–609.

42. Kirkham PM, Schroeder HW Jr. Antibody structure and the evolution of immunoglobulin V gene segments. *Semin Immunol* 1994;6:347–360.

43. Kirkham PM, Elgavish RA, Schroeder HW Jr. Structure and evolution of mammalian $V_κ$ families. *J Immunol* 1993;150:151a.

44. Kroemer G, Helmberg A, Bernot A, Auffray C, Kofler R. Evolutionary relationship between human and mouse immunoglobulin κ light chain variable region genes. *Immunogenetics* 1991;33:42–49.

45. Pascual V, Capra JD. Human immunoglobulin heavy-chain variable region genes: organization, polymorphism, and expression. *Adv Immunol* 1991;49: 1–74.

46. Zachau HG. The immunoglobulin κ locus—or—what has been learned from looking closely at one-tenth of a percent of the human genome. *Gene* 1993;135: 167–173.

47. Frippiat JP, Williams SC, Tomlinson IM, et al. Organization of the human immunoglobulin λ light-chain locus on chromosome 22q11.2. *Hum Mol Genet* 1995;4(6):983–991.

48. Cook GP, Tomlinson IM, Walter G, et al. A map of the human immunoglobulin V_H locus completed by analysis of the telomeric region of chromosome 14q. *Nat Genet* 1994;7(2):162–168.

49. Matsuda F, Shin EK, Nagaoka H, et al. Structure and physical map of 64 variable segments in the 3' 0.8 megabase region of the human immunoglobulin heavy chain locus. *Nat Genet* 1993;3:88–94.

50. Matsuda F, Honjo T. Organization of the human immunoglobulin heavy chain locus. *Adv Immunol* 1996;62:1–29.

51. Hsu E. The variation in immunoglobulin heavy chain C regions in evolution. *Semin Immunol* 1994;6:383–391.

52. Davis AC, Roux KH, Pursey J, Shulman MJ. Intermolecular disulfide bonding in IgM: effects of replacing cysteine residues in the μ heavy chain. *EMBO J* 1989;8:25–19–2526.

53. Mestecky J, Russell MW. IgA subclasses. *Monogr Allergy* 1986;19:277–301.

54. Burnett RC, Hanly WC, Zhai SK, Knight KL. The IgA heavy chain gene family in rabbit: cloning and sequence analysis of 13 Cα genes. *EMBO J* 1989;8: 4041–4047.

55. Poljak RJ. X-ray diffraction studies of immunoglobulins. *Adv Immunol* 1975;21: 1–33.

56. Huck S, Fort P, Crawford DH, Lefranc MP, Lefranc G. Sequence of a human IgG3 heavy chain C region gene: comparison with the other human Cγ genes. *Nucleic Acids Res* 1986;14:1779–1789.

57. Kocher HP, Bijlenga RK, Jaton JC. Biosynthesis and structure of membrane and secretory immunoglobulins. *Mol Cell Biochem* 1982;47:11–22.

58. Rogers J, Choi E, Souza L, et al. Gene segments encoding transmembrane carboxyl termini of Ig γ chains. *Cell* 1981:19–27.

58a. Tomlinson IM, Walter G, Jones PT, Dear PH, Sonnhammer ELL, Winter G. The imprint of somatic hypermutation on the repertoire of human germline V genes. *J Mol Biol* 1996;256:813–817.

59. Davies DR, Padlan EA, Sheriff S. Antigen-antibody complexes. *Annu Rev Biochem* 1990;59:439–473.

60. Wright A, Tao MH, Kabat EA, Morrison SL. Antibody variable region glycosylation: position effects on antigen binding and carbohydrate structure. *EMBO J* 1991;10:2717–2723.

61. Middaugh CR, Litman GW. Atypical glycosylation of an IgG monoclonal cryo-Ig. *J Biol Chem* 1987;262:3671–3673.

62. Chothia C, Lesk AM, Tramontano A, et al. Conformations of immunoglobulin hypervariable regions. *Nature* 1989;342:877–883.

63. Strong RK, Campbell R, Rose DR, Petsko GA, Sharon J, Margolies MN. Three-dimensional structure of murine anti-p-azophenylarsonate Fab 36–71. *Biochemistry* 1991;30:3739–3748.

64. Guddat LW, Herron JN, Edmundson AB. Three-dimensional structure of a

65. Padlan EA, Davies DR. A model of the Fc of IgE. *Mol Immunol* 1986;10: 1063–1075.

66. Carayannopoulos L, Hexham JM, Capra JD. Localization of the binding site for the monocyte IgA-Fc receptor (CD89) to the domain boundary between $C_α2$ and $C_α3$ in human IgA1. *J Exp Med* 1996;183:1579–1586.

66a. Putnam FW. *The plasma proteins: structure, function, and genetic control,* vol 3, 2nd ed. New York: Academic Press, 1977.

67. Grey HM, Abel CA, Yount WJ, Kunkel HG. A subclass of human γA-globulins (γ-A2) which lacks the disulfide bonds linking heavy and light chains. *J Exp Med* 1968;128:1223–1236.

68. Saul FA, Amzel LM, Poljak RJ. Preliminary refinement and structural analysis of the Fab fragment from human immunoglobulin New at 2.0 A resolution. *J Biol Chem* 1978;253:585–597.

69. Padlan EA. Anatomy of the antibody molecule. *Mol Immunol* 1994;31: 169–217.

70. Davies DR, Sheriff S, Padlan E. Comparative study of two Fab-lysozyme crystal structures. *Cold Spring Harb Symp Quant Biol* 1989;54:233–238.

71. Schumaker VN, Phillips ML, Hansen DC. Dynamic aspects of antibody structure. *Mol Immunol* 1991;28:1347–1360.

72. Wade RH, Tavequ JC, Lamy JN. Concerning the axial flexibility of the Fab regions of IgG. *J Mol Biol* 1989;206:349–356.

73. Dangl JL, Wensel TG, Morrison SL, Stryer L, Herzenberg LA, Oi VT. Segmental flexibility and complement fixation of genetically engineered chimeric human, rabbit and mouse Abs. *EMBO J* 1988;7:1989–1994.

74. Spiegelberg HL. The structure and biology of human IgD. *Immunol Rev* 1977; 37:3–23.

75. Putnam FW, Liu YS, Low TL. Primary structure of a human IgA1 immunoglobulin. *J Biol Chem* 1979;254:2865–2874.

76. Plaut AG, Wistar R Jr, Capra JD. Differential susceptibility of human IgA immunoglobulins to streptococcal IgA protease. *J Clin Invest* 1974;54: 1295–1300.

77. Plaut AG, Gilbert JV, Artenstein MS, Capra JD. *Neisseria gonorrhoeae* and *Neisseria meningitidis:* extracellular enzyme cleaves human IgA. *Science* 1975;190: 1103–1105.

78. Dwek R. Glycobiology—towards understanding the function of sugars. *Biochem Soc Trans* 1995;23:1–25.

79. Jenkins M, Curling E. Glycosylation of recombinant proteins: problems and prospects. *Enzyme Microb Technol* 1995;16:354–364.

80. Schaffner G, Haase M, Geiss S. Criteria for investigation of the product equivalence of monoclonal antibodies for therapeutic and *in vivo* diagnostic use in case of change in manufacturing process. *Biologicals* 1995;23:253–259.

81. Beale D, Feinstein A. Structure and function of the C regions of immunoglobulins. *Q Rev Biophys* 1976;9:135–180.

81a. Jefferis R, Lund J. Glycosylation of antibody molecules: structural and functional significance. *Chem Immunol* 1997;65:111–128.

82. Tsuchiya N, Endo T, Matsuta K, et al. Effects of galactose depletion from oligosaccharide chains on immunological activities of human IgG. *J Rheumatol* 1989;16:285–290.

83. Tao MH, Morrison SL. Studies of aglycosylated chimeric mouse-human IgG. *J Immunol* 1989;143:2595–2601.

84. Wright A, Morrison SL. Effect of altered C_H2-associated carbohydrate structure on the functional properties and *in vivo* fate of human IgG1. *J Exp Med* 1994; 180:1087–1096.

85. Lund J, Winter G, Jones PT, et al. Human FcγRI and FcγRII interact with distinct but overlapping binding sites on the human IgG. *J Immunol* 1991;147: 2657–2662.

86. Canfield SM, Morrison SL. The binding affinity of human IgG for its high affinity Fc receptor is determined by multiple amino acids in the C_H2 domain and is modulated by the hinge region. *J Exp Med* 1991;173:1483–1491.

87. Jefferis R, Lund J. Goodall M. Recognition sites on human IgG for Fcγ receptors: the role of glycosylation. *Immunol Lett* 1995;44:111–117.

87a. Davis AC, Roux KH, Shulman MJ. On the structure of polymeric IgM. *Eur J Immunol* 1988;18:1001–1008.

88. Mole JE, Bhown AS, Bennett JC. Primary structure of human J chain: alignment of peptides from chemical and enzymatic hydrolyses. *Biochemistry* 1977;16: 3507–3513.

89. Koshland ME. The coming of age of the immunoglobulin J chain. *Annu Rev Immunol* 1985;3:425–453.

90. Max EE, Korsmeyer SJ. Human J chain gene. *J Exp Med* 1985;161:832–849.

91. Cann GM, Zaritsky A, Koshland ME. Primary structure of the immunoglobulin J chain from the mouse. *Proc Natl Acad Sci USA* 1982;79:6656–6660.

92. Hughes GJ, Frutiger S, Paquet N, Jaton JC. The amino acid sequence of rabbit J chain in secretory IgA. *Biochem J* 1990;271:641–647.

93. Mikoryak CA, Margolies MN, Steiner LA. J chain in *Rana catesbeiana* high molecular weight Ig. *J Immunol* 1988;140:4279–4285.

94. Takahashi T, Iwase T, Takenouchi N, et al. The joining (J) chain is present in invertebrates that do not express immunoglobulins. *Proc Natl Acad Sci USA* 1996;93:1886–1891.

95. Zikan J, Novotny J, Trapane TL, et al. Secondary structure of the immunoglobulin J chain. *Proc Natl Acad Sci USA* 1985;82:5905–5909.

96. Davis AC, Collins C, Yoshimura MI, D'Agostaro G, Shulman MJ. Mutations of

human immunoglobulin with a hinge deletion. *Proc Natl Acad Sci USA* 1993;90:4271–4275.

the mouse μ chain which prevent polymer assembly. *J Immunol* 1989;143: 1352–1357.

97. Mather EL, Alt FW, Bothwell A, Baltimore D, Koshland ME. Expression of J chain RNA in cell lines representing different stages of B lymphocyte differentiation. *Cell* 1981;23:369–378.

98. Randall TD, Parkhouse RM, Corley RB. J chain synthesis and secretion of hexameric IgM is differentially regulated by LPS and IL5. *Proc Natl Acad Sci USA* 1992;89:962–966.

99. Randall TD, King LB, Corley RB. The biological effects of IgM hexamer formation. *Eur J Immunol* 1990;20:1971–1979.

100. Underdown B. Transcytosis by the receptor for polymeric immunoglobulin. In: Metzger H, ed. *Fc receptors and the action of antibodies.* Washington, DC: American Society for Microbiology, 1990.

101. Mostov KE, Blobel G. A transmembrane precursor of secretory component. *J Biol Chem* 1982;257:11816–11821.

102. Mostov KE, Friedlander M, Blobel G. The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains. *Nature* 1984; 308:37–43.

103. Brandtzaeg P. Human secretory IgM. *Immunology* 1975;29:559–570.

104. Weicker J, Underdown BJ. A study of the association of human secretory component with IgA and IgM proteins. *J Immunol* 1975;114:1337–1344.

105. Mach J-P. *In vitro* combination of human and bovine free secretory component with IgA of various species. *Nature* 1970;228:1278–1282.

106. Bienenstock J, Perey DY, Gauldie J, Underdown BJ. Chicken γA: physiochemical and immunochemical characteristics. *J Immunol* 1973;110:524–533.

107. Fallgreen-Gebauer E, Bebauer W, Bastian A, et al. The covalent linkage of secretory component to IgA. *Biol Chem Hoppe Seyler* 1993;374:1023–1028.

108. Bakos MA, Kurosky A, Goldblum RM. Characterization of a critical binding site for human polymeric immunoglobulin on secretory component. *J Immunol* 1991;147:3419–3426.

109. Hasemann C, Capra JD. Mutational analysis of arsonate binding by a CRIA+ antibody. *J Biol Chem* 1991;266:7626–7632.

110. Amit AG, Mariuzza RA, Phillips SEV, Poljak RJ. Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science* 1986;233:747–753.

110a.Davies DR, Sheriff S, Padlan EA, Silverton EW, Cohen GH, Smith-Gill SJ. Three-dimensional structures of two Fab complexes with lysozyme. In: Smith-Gill SJ, Sercarz EE, eds. *The immune response to structurally define proteins: the lysozyme model.* Schenectady, NY: Adenine Press, 1989.

111. Padlan EA. On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins* 1990;7:112–124.

112. Mian IS, Bradwell AR, Olson AJ. Structure, function and properties of antibody binding sites. *J Mol Biol* 1991;217:133–151.

113. Mariuzza RA, Poljak RJ, Schwarz FP. The energetics of antigen-antibody binding. *Res Immunol* 1994;145:70–72.

114. Braden BC, Poljak RJ. Structural features of the reactions between antibodies and protein antigens. *FASEB J* 1995;9:9–16.

115. Herman A, Kappler JW, Marrack P, Pullen AM. Superantigens: mechanism of T-cell stimulation and role in immune responses. *Annu Rev Immunol* 1991;9:745–772.

116. Pascual V, Capra JD. B-cell superantigens? *Curr Biol* 1991;1:315.

117. Silverman GJ. Superantigens and spectrum of unconventional B-cell antigens. *The Immunologist* 1994;2:51–57.

118. Biguzzi S. Fcγ-like determinants on immunoglobulin variable regions: identification by staphylococcal protein A. *Scand J Immunol* 1982;15:605–618.

119. Sasso EH, Silverman GJ, Mannik M. Human IgM molecules that bind staphylococcal protein A contain V$_H$III H chains. *J Immunol* 1989;142:2778–2783.

120. Potter KN, Li Y, Capra JD. Staphylococcal protein A simultaneously interacts with FR1, CDR2, and FR3 on human V$_H$3-encoded Igs. *J Immunol* 1996;157: 2982–2988.

121. Berberian L, Goodglick L, Kipps TJ, Braun J. Immunoglobulin V$_H$3 gene products: natural ligands for HIV gp120. *Science* 1993;261:1588–1591.

122. Lenert P, Kroon D, Spiegleberg H, Golub ES, Zanetti M. Human CD4 binds immunoglobulins. *Science* 1990;248:1639–1643.

123. Corper AL, Sohi MK, Bonagura VR, et al. Structure of human IgM rheumatoid factor Fab bound to its autoantigen IgG Fc reveals a novel topology of antibody-antigen interaction. *Nat Struct Biol* 1997;4:374–381.

123a.Simard J, Mak TW. Immunoglobulins and the B cell receptor. In: *Genetic basis of immune responses* (in press).

124. Hombach J, Lottspeich F, Reth M. Identification of the genes encoding the IgM-α and Ig-β components of the IgM antigen receptor complex by amino-terminal sequencing. *Eur J Immunol* 1990;20:2795–2799.

125. Williams GT, Venkitaraman AR, Gilmore DJ, Neuberger MS. The sequence of the μ transmembrane segment determines the tissue specificity of the transport of IgM to the cell surface. *J Exp Med* 1990;171:947–952.

126. Karasuyama H, Kudo A, Melchers F. The proteins encoded by the VpreB and λ5 pre-B cell-specific genes can associate with each other and with μ heavy chain. *J Exp Med* 1990;172:969–972.

127. Tsubata T, Reth M. The products of pre-B cell-specific genes (λ5 and VpreB) and the immunoglobulin μ chain form a complex that is transported onto the cell surface. *J Exp Med* 1990;172:973–976.

128. Arya S, Chen F, Spycher S, Isenman DE, Shulman MJ, Painter RH. Mapping of amino acid residues in the C$_μ$3 domain of mouse IgM important in macromole-

cular assembly and complement-dependent cytolysis. *J Immunol* 1994;152: 1206–1212.

129. Wright JF, Shulman MJ, Isenman DE, Painter RH. C1 binding by mouse IgM. *J Biol Chem* 1990;265:10506–10513.

130. Du Pasquier L, Wabl MR. Antibody diversity in amphibians: inheritance of isoelectric focusing antibody patterns in isogenic frogs. *Eur J Immunol* 1978;8: 428–433.

131. Lin LC, Putnam FW. Primary structure of the Fc region of human IgD: implications of evolutionary origin and biologic function. *Proc Natl Acad Sci USA* 1981;78:504–508.

132. Nitschke L, Kosco MH, Kohler G, Lamers MC. IgD-deficient mice can mount normal immune responses to thymus-independent and -dependent antigens. *Proc Natl Acad Sci USA* 1993;90:1887–1891.

133. Roes J, Rajewsky K. Immunoglobulin D (IgD)-deficient mice reveal an auxiliary receptor function for IgD in antigen-mediated recruitment of B cells. *J Exp Med* 1993;177:45–55.

134. Forster I, Vieira P, Rajewsky K. Flow cytometric analysis of cell proliferation dynamics in the B cell compartment of the mouse. *Int Immunol* 1989;1(4): 321–331.

135. Black SJ, van der Loo W, Loken MR, Herzenberg LA. Expression of IgD by murine lymphocytes. Loss of surface IgD indicates maturation of memory B cells. *J Exp Med* 1978;147:984–996.

136. McHeyzer-Williams MG, Nossal GJ, Lalor PA. Molecular characterization of single memory B cells. *Nature* 1991;350:502–505.

137. Havran WL, DiGiusto DL, Cambier JC. mIgM:mIgD ratios on B cells: mean mIgD expression exceeds mIgM by 10-fold on most splenic B cells. *J Immunol* 1984;132:1712–1716.

138. Venkitaraman AR, Williams GT, Dariavach P, Neuberger MS. The B-cell antigen receptor of the five immunoglobulin classes. *Nature* 1991;352:777–781.

139. Nezlin R. Internal movements in immunoglobulin molecules. *Adv Immunol* 1990;48:1–40.

140. Simard J, Mak TW. Immunoglobulins and the B cell receptor. In: *Genetic basis of immune responses* (in press).

141. Brink R, Goodnow CC, Crosbie J, et al. IgM and IgD antigen receptors are both capable of mediating B lymphocyte activation, deletion, or anergy after interaction with specific antigen. *J Exp Med* 1992;176:991–1005.

142. Morris SC, Lees A, Finkelman FD. *In vivo* activation of naïve T cells by antigen-presenting B cells. *J Immunol* 1994;152:3777–3785.

143. Brink R, Goodnow CC, Basten A. IgD expression on B cells is more efficient than IgM but both receptors are functionally equivalent in up-regulating CD80/CD86 co-stimulatory molecules. *Eur J Immunol* 1995;25:1980–1984.

144. Swenson CD, Van Vollenhoven RF, Xue B, Siskind GW, Thorbecke GJ, Coico RF. Physiology of IgD. *Eur J Immunol* 1988;18:13–20.

145. Finkelman FD, Snapper CM, Mountz JD, Katona IM. Polyclonal activation of the murine immune system by a goat antibody to mouse IgD. *J Immunol* 1987;138:2826–2830.

146. Liu Y-J, de Bouteiller O, Arpin C, et al. Normal human IgD+/IgM+ germinal center B cells can express up to 80 mutations in the variable region of their IgD transcripts. *Immunity* 1996;4:603–613.

147. Billian G, Bella C, Mondiere P, Defrance T. Identification of a tonsil IgD+ B cell subset with phenotypical and functional characteristics of germinal center B cells. *Eur J Immunol* 1996;26:1712–1719.

148. Perlmutter RG, Hansburg D, Briles DE, Nicolotti RA, Davis JM. Subclass restriction of murine anti-carbohydrate antibodies. *J Immunol* 1978;121: 566–571.

149. Coutelier J-P, van der Logt JTM, Heessen FWA, Warnier G, Van Snick J. IgG2a restriction of murine antibodies elicited by viral infections. *J Exp Med* 1987;165: 64–69.

150. Papadea C, Check IJ. Human IgG and IgG subclasses: biochemical, genetic, and clinical aspects. *Crit Rev Clin Lab Sci* 1989;27:27–58.

151. Brekke OH, Michaelsen TE, Sandlie I. The structural requirements for complement activation by IgG: does it hinge on the hinge? *Immunol Today* 1995;16: 85–90.

152. Bruggemann M, Williams GT, Bindon CI, et al. Comparison of the effector functions of human immunoglobulins using a matched set of chimeric antibodies. *J Exp Med* 1987;166:1351–1361.

153. Duncan AR, Winter G. The binding site for C1q on IgG. *Nature.* 1988;332: 738–740.

154. Tao M-H, Canfield SM, Morrison SL. The differential ability of human IgG1 and IgG4 to activate complement is determined by the COOH-terminal sequence of the C$_H$2 domain. *J Exp Med* 1991;173:1025–1028.

155. Tan LK, Shopes RJ, Oi VT, Morrison SL. Influence of the hinge region on complement activation, C1q binding, and segmental flexibility in chimeric human immunoglobulins. *Proc Natl Acad Sci USA* 1990;87:162–166.

156. Sandlie I, Michaelsen TE. Engineering monoclonal antibodies to determine the structural requirements for complement activation and complement mediated lysis. *Mol Immunol* 1991;28:1361–1368.

157. Brekke OH, Michaelsen TE, Sandin R, Sandlie I. Activation of complement by an IgG molecule without a genetic hinge. *Nature* 1993;363:628–630.

158. Norderhaug L, Brekke OH, Bremnes B, et al. Chimeric mouse-human IgG3 antibodies with an IgG4-like hinge region induce complement-mediated lysis more efficiently than IgG3 with normal hinge. *Eur J Immunol* 1991;21:2379–2384.

159. Kimberly RP, Salmon JE, Edberg JC. Receptors for IgG. *Arthritis Rheum* 1995;38:306–314.

160. Indik ZK, Park JG, Hunter S, Schreiber AD. Structure/function relationships of Fcγ receptors in phagocytosis. *Semin Immunol* 1995;7:45–54.

161. Unkeless JC. Function and heterogeneity of human Fc receptors for IgG. *J Clin Invest* 1989;83:355–361.

162. Michaelsen TE, Aase A, Norderhaug L, Sandlie I. Antibody dependent cell-mediated cytotoxicity induced by chimeric mouse-human IgG subclasses and IgG3 antibodies with altered hinge region. *Mol Immunol* 1992;29:319–326.

163. Galon J, Bouchard C, Fridman WH, Sautes C. Ligands and biological activities of soluble Fcγ receptors. *Immunol Lett* 1995;44:175–181.

164. Saji F, Koyama M, Matsuzaki N. Current topic: human placental Fc receptors. *Placenta* 1994;15:453–466.

165. Sarmay G, Lund J, Rozsnyay Z, Gergely J, Jefferis R. Mapping and comparison of the interaction sites on the Fc region of IgG responsible for triggering ADCC through different types of human Fcγ receptor. *Mol Immunol* 1992;29:633–639.

166. Canfield SM, Morrison SL. The binding affinity of human IgG for its high affinity Fc receptor is determined by multiple amino acids in the C_H2 domain and is modulated by the hinge region. *J Exp Med* 1991;173:1483–1491.

167. Tomasi TB, Grey HM. Structure and function of IgA. *Prog Allergy* 1972;16:81–213.

168. Lehner T, Bergmeier LA, Tao L, et al. Targeted lymph node immunization with simian immunodeficiency virus p27 antigen to elicit genital, rectal, and urinary immune responses in nonhuman primates. *J Immunol* 1994;153:1858–1868.

169. Mestecky J, McGhee JR. Immunoglobulin A (IgA): molecular and cellular interactions involved in IgA biosynthesis and immune response. *Adv Immunol* 1987;40:153–245.

170. Qiu JZ, Brackee GP, Plaut AG, Sneller MC, Strober W. Analysis of the specificity of bacterial IgA proteases by a comparative study of ape serum IgAs as substrates. *Infect Immun* 1996;64:933–937.

171. Wang B, Rieger A, Kilgus O, et al. Human epidermal Langerhans cells from normal human skin bind monomeric IgE via FcεRI. *J Exp Med* 1992;175:1353–1365.

172. Bieber T, de la Salle H, Wollenberg A, et al. Human epidermal Langerhans cells express the high affinity receptor for IgE (FcεRI). *J Exp Med* 1992;175:1285–1290.

173. Gounni AS, Lamkhioued B, Ochiai K, et al. High-affinity IgE receptor on eosinophils is involved in defense against parasites. *Nature* 1994;367:183–186.

174. Hashimoto S, Koh K, Tomita Y, et al. TNF-α regulates IL-4-induced FcεRII/CD23 gene expression and soluble FcεRII release by human monocytes. *Int Immunol* 1995;7:705–713.

175. Helm B, Marsh P, Vercelli D, Padlan E, Gould H, Geha R. The mast cell binding site on human immunoglobulin E. *Nature* 1988;331:180–183.

176. Nissim. A, Eshhar Z. The human mast cell receptor binding site maps to the third constant domain of IgE. *Mol Immunol* 1992;29:1065–1072.

177. Edelman GM, Cunningham BA, Gall WE, Gottlieb PD, Rutishauser U, Waxdal MJ. The covalent structure of an entire γG immunoglobulin molecule. *Proc Natl Acad Sci USA* 1969;63:78–85.

178. Brummendorf T, Rathjen FG. Structure/function relationships of axon-associated adhesion receptors of the immunoglobulin superfamily. *Curr Opin Neurobiol* 1996;6:584–593.

179. Edelman GM. CAMs and Igs: cell adhesion and the evolutionary origins of immunity. *Immunol Rev* 1987;100:11–45.

180. Hultgren SJ, Jacob-Dubuisson F, Jones CH, Branden CI. PapD and superfamily of periplasmic immunoglobulin-like pilus chaperones. *Adv Protein Chem* 1993;44:99–123.

181. Bateman A, Eddy SR, Chothia C. Members of the immunoglobulin superfamily in bacteria. *Protein Sci* 1996;5:1939–1941.

182. Wojciechowicz D, Lu CF, Kurjan J, Lipke PN. Cell surface anchorage and ligand-binding domains of the *Saccharomyces cerevisiae* cell adhesion protein alpha-agglutinin, a member of the IgSF. *Mol Cell Biol* 1993;13:2554–2563.

183. Ohno S. The origin of Igs and T-cell receptors is likely to be the cell death sensor of macrophages. *Res Immunol* 1996;147:247–252.

184. Noegel A, Gerisch G, Stadler J, Westfal M. Complete sequence and transcript regulation of a cell adhesion protein from aggregating *Dictyostelium* cells. *Embo J* 1986;5:1473–1476.

185. Mokady O. Occam's Razor, invertebrate allorecognition and immunoglobulin superfamily evolution. *Res Immunol* 1996;147:241–246.

186. Lesk AM, Chothia C. Evolution of proteins formed by β-sheets. *J Mol Biol* 1982;160:325–342.

187. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J* 1986;5:823–826.

188. Harpaz Y, Chothia C. Many of the IgSF domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J Mol Biol* 1994;238:528–539.

189. Kolsch E, Oberbarnscheidt J, Bruner K, Heuer J. The Fc receptor: its role in the transmission of differentiation signals. *Immunol Rev* 1980;49:61–78.

190. Fridman WM. Fc receptors and immunoglobulin binding factors. *FASEB J* 1991;5:2684–2690.

191. Unkeless JC, Scigliano, E, Freedman VH. Structure and function of human and murine receptors for IgG. *Annu Rev Immunol* 1988;6:251–281.

192. Sandor M, Lynch R. The biology and pathology of Fc receptors. *J Clin Immunol* 1993;13:237–246.

193. Keegan AD, Paul WE. Multichain immune recognition receptors: similarities in structure and signaling pathways. *Immunol Today* 1992;13:63–68.

194. Lanier L, Yu G, Phillips J. Analysis of FcγRIII (CD16) membrane expression and association with CD3ζ and FcεRI-γ by site directed mutation. *J Immunol* 1991;146:1571–1576.

195. Cassel DL, Keller MA, Surrey S, et al. Differential expression of FcγRIIA, FcγRIIB and FcγRIIC in hematopoietic cells: analysis of transcripts. *Mol Immunol* 1993;30:451–460.

196. Heyman B. The immune complex: possible ways of regulating the antibody response. *Immunol Today* 1990;11:310–313.

197. Vivier E, Daeron M. Immunoreceptor tyrosine-based inhibition motifs. *Immunol Today* 1997;18:286–291.

198. Hogarth PM, Hulett MD, Ierino FL, Tate B, Powell MS, Brinkworth RL. Identification of the immunoglobulin binding regions of FcγRII and FcεRI. *Immunol Rev* 1992;125:21–35.

199. Hulett MD, Witort E, Brinkworth RI, McKenzie IF, Hogarth PM. Multiple regions of human FcγRII (CD32) contribute to the binding of IgG. *J Biol Chem* 1995;270:21188–21194.

200. Letourneur O, Kennedy IC, Brini AT, Ortaldo JR, O'Shea JJ, Kinet JP. Characterization of the family of dimers associated with Fc receptors (FcεRI and FcγRIII). *J Immunol* 1991;147:2652–2656.

201. Selvaraj P, Rosse WF, Silber R, Springer TA. The major Fc receptor in blood has a phosphatidylinositol anchor and is deficient in paroxysmal nocturnal haemoglobinuria. *Nature* 1988;333:565–567.

202. Metzger H. The receptor with high affinity for IgE. *Immunol Rev* 1992;125:38–48.

203. Maurer D, Fiebiger E, Reininger B, et al. Expression of functional high affinity IgE receptors (FcεRI) on monocytes of atopic individuals. *J Exp Med* 1994;179:745–750.

204. Dombrowicz D, Flamand V, Brigman K, et al. Abolition of anaphylaxis by targeted disruption of the high affinity IgE receptor alpha chain gene. *Cell* 1993;75:969–976.

205. Takai T, Li M, Sylvestre D, et al. FcR γ chain deletion results in pleiotropic effector cell defects. *Cell* 1994;76:519–529.

206. Clynes R, Ravetch J. Cytotoxic antibodies trigger inflammation through Fc receptors. *Immunity* 1995;3:21–26.

207. Leahy DJ. A structural view of CD4 and CD8. *FASEB J* 1995;9:17–25.

208. Rosenstein Y, Ratnofshy S, Burakoff SJ, Herrmann SH. Direct evidence for binding of CD8 to HLA class I antigens. *J Exp Med* 1989;169:149–160.

209. Weber S, Karjalainen K. Mouse CD4 binds MHC class II with extremely low affinity. *Int Immunol* 1993;5:695–698.

210. Littman DR. The structure of the CD4 and CD8 genes. *Annu Rev Immunol* 1987;5:561–584.

211. Veillette A, Bookman MA, Horak EM, Bolen JB. The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine-protein kinase p56^{lck}. *Cell* 1988;55:301–308.

212. Wallace VA, Penninger J, Mak TW. CD4, CD8 and tyrosine kinases in thymic selection. *Curr Opin Immunol* 1983;5:235–240.

213. Janeway CA. The T cell receptor as a multicomponent signaling machine: CD4/CD8 coreceptors, and CD45 in T cell activation. *Annu Rev Immunol* 1992;10:645–674.

214. Littman DR, Thomas Y, Maddon PJ, Chess L, Axel R. The isolation and sequence of the gene encoding T8: a molecule defining functional classes of T lymphocytes. *Cell* 1985;40:237–246.

215. Norment AM, Littman DR. A second subunit of CD8 is expressed in human T cells. *EMBO J* 1988;7:3433–3439.

216. Shiue L, Gorman SD, Parnes JR. A second chain of human CD8 is expressed on peripheral blood lymphocytes. *J Exp Med* 1988;168:1993–2005.

217. Wheeler CJ, von Hoegen P, Parnes JR. An immunological role for the CD8β-chain. *Nature* 1992;357:247–249.

218. Nakayama K, Nakayama K, Negishi I, et al. Requirement for CD8β chain in positive selection of CD8-lineage T cells. *Science* 1994;263:1131–1133.

219. Renard V, Romero P, Vivier E, Malissen B, Luescher IF. CD8β increases CD8 coreceptor function and participation in TCR-ligand binding. *J Exp Med* 1996;184:2439–2444.

220. Irie HY, Ravichandran KS, Burakoff SJ. CD8β chain influences CD8α chain-associated Lck kinase activity. *J Exp Med* 1995;181:1267–1273.

221. Nakayama K, Kawachi Y, Tokito S, et al. Recent duplication of the two human CD8 β-chain genes. *J Immunol* 1992;148:1919–1927.

222. DiSanto JP, Smith D, de Bruin D, Lacy E, Flomenberg N. Transcriptional diversity at the duplicated human CD8β loci. *Eur J Immunol* 1993;23:320–326.

223. Leahy DJ, Axel R, Hendrickson WA. Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 Å resolution. *Cell* 1992;68:1145–1162.

224. Kirszbaum L, Sharpe JA, Goss N, Lahnstein J, Walker ID. The α-chain of murine CD8 lacks an invariant Ig-like disulfide bond but contains a unique intrachain loop instead. *J Immunol* 1989;142:3931–3936.

225. Classon BJ, Brown MH, Garnett D, et al. The hinge region of the CD8α chain: structure, antigenicity, and utility in expression of IgSF domains. *Int Immunol* 1992;4:215–225.

226. Pascale MC, Erra MC, Malagolini N, Serafini-Cessi F, Leone A, Bonatti S. Post-translational processing of an O-glycosylated protein, the human CD8 glycopro-

tein, during the intracellular transport to the plasma membrane. *J Biol Chem* 1992;267:25196–25201.

227. Boursier JP, Alcover A, Herve F, Laisney I, Acuto O. Evidence for an extended structure of the T-cell co-receptor CD8α as deduced from the hydrodynamic properties of soluble forms of the extracellular region. *J Biol Chem* 1993;268:2013–2020.

228. Sanders SK, Fox RO, Kavathas P. Mutations in CD8 that affect interactions with HLA class I and monoclonal anti-CD8 antibodies. *J Exp Med* 1991;174:371–379.

229. Giblin PA, Leahy DJ, Mennone J, Kavathas PB. The role of charge and multiple faces of the CD8 α/α homodimer in binding to MHC class I molecules: support for a bivalent model. *Proc Natl Acad Sci USA* 1994;91:1716–1720.

230. Gao GF, Tormo J, Gerth UC, et al. Crystal structure of the complex between human CD8α/α and HLA-A2. *Nature* 1997;387:630–634.

231. Maddon PJ, Littman DR, Godfrey M, Maddon DE, Chess L, Axel R. The isolation and nucleotide sequence of a cDNA encoding the T-cell surface protein T4: a new member of the IgSF. *Cell* 1985;42:93–104.

232. Wang JH, Yau YW, Garrett TP, et al. Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature* 1990;348:411–418.

233. Ryu SE, Kwong PD, Truneh A, et al. Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* 1990;348:419–426.

234. Brady RL, Dodson EJ, Dodson GG, et al. Crystal structure of domains 3 and 4 of rat CD4: relation to the NH2-terminal domains. *Science* 1993;260:979–983.

235. Lange G, Lewis SJ, Murshudov GN, et al. Crystal structure of an extracellular fragment of the rat CD4 receptor containing domains 3 and 4. *Structure* 1994;2: 469–481.

236. Williams AF, Davis SJ, He Q, Barclay AN. Structural diversity in domains of the immunoglobulin superfamily. *Cold Spring Harb Symp Quant Biol* 1989;54: 637–647.

237. Wu H, Kwong PD, Hendrickson WA. Dimeric association and segmental variability in the structure of human CD4. *Nature* 1997;387:527–530.

238. Fleury S, Lamarre D, Meloche S, et al. Mutational analysis of the interaction between CD4 and class II MHC: class II antigens contact CD4 on a surface opposite the gp120-binding site. *Cell* 1991;66:1037–1049.

239. Moebius U, Pallai P, Harrison SC, Reinherz EL. Delineation of an extended surface contact area on human CD4 involved in class II major histocompatibility complex binding. *Proc Natl Acad Sci USA* 1993;90:8259–8263.

240. Huang B, Yachou A, Fleury S, Hendrickson WA, Sekaly RP. Analysis of the contact sites on the CD4 molecule with class II MHC molecule: co-ligand versus coreceptor function. *J Immunol* 1997;158:216–225.

241. Langedijk JP, Puijk WC, van Hoorn WP, Meloen RH. Location of CD4 dimerization site explains critical role of CDR3-like region in HIV-1 infection and T-cell activation and implies a model for complex of coreceptor-MHC. *J Biol Chem* 1993;268:16875–16878.

Appendix D

AN ANALYSIS OF THE SEQUENCES OF THE VARIABLE REGIONS OF BENCE JONES PROTEINS AND MYELOMA LIGHT CHAINS AND THEIR IMPLICATIONS FOR ANTIBODY COMPLEMENTARITY*

By TAI TE WU, Ph.D., and ELVIN A. KABAT, Ph.D

(*From the Departments of Microbiology, Neurology, and Human Genetics and Development, College of Physicians and Surgeons, Columbia University, and the Neurological Institute, Presbyterian Hospital, New York 10032; the Biomathematics Division, Graduate School of Medical Sciences, Cornell University and the Sloan-Kettering Institute, New York 10021*)

The extraordinary versatility of the antibody-forming mechanism in producing an almost limitless number of specific receptor sites complementary for almost any molecular conformation of matter within a size range (1–3) represented by a hexa- or heptasaccharide as an upper and a mono- or disaccharide as a lower limit, is almost certainly related to the unique structural features of immunoglobulins and differentiates them from all other known proteins. These antibody-combining sites are formed as a consequence of the interaction of two polypeptide chains, a light and a heavy chain (2, 4, 5). The antibodies usually formed to various antigens often represent heterogeneous populations of immunoglobulin molecules of different classes, subclasses, and genetic variants and also show specificities toward different antigenic determinants (1, 2, 6, 7). In some instances, however, relatively homogeneous populations of antibodies with respect to many of these properties have been obtained. Among these have been human antibodies to dextran and levan (8, 9) and rabbit antibodies to the group-specific carbohydrate of streptococcus (10–12), antibodies to the Type III–specific capsular polysaccharide of pneumococcus (13, 14), rabbit antihapten (15), and specimens of antibodies and of Fab' fragments which crystallized (Nisonoff et al., in references 16, 17), but sequence data on these are not yet available.

The large body of sequence data related to immunoglobulin structure comes from the analysis of urinary Bence Jones proteins and from the monoclonal immunoglobulins found in large amounts in the sera of patients with multiple myeloma and Waldenström macroglobulinemia (16, 18). While a substantial body of evidence was available relating these proteins to immunoglobulins, the recent demonstration that many myeloma globulins have specific ligand-binding properties like those of many antibodies provides increasing confidence that myeloma globulins represent homogeneous populations of antibody molecules (16, 18–27). The ability to produce in BALB/c

211

mice myelomas and macroglobulinemias (28) which produce myeloma globulins and Bence Jones proteins like those in the human, provides a source of data from which important evolutionary trends can be inferred.

Thus the extensive sequence data on Bence Jones proteins, which are considered to be light chains of myeloma globulins and Waldenström macroglobulins (29), and on various light and heavy chains, provide information clearly pertinent to the problem of the elucidation of the structure of antibody-combining sites.

The unique finding that distinguishes the immunoglobulins from all other proteins is that the N-terminal half of the light chains and the N-terminal quarter of the heavy chains vary in sequence in samples obtained from individual monoclonal immunoglobulins and that indeed no two such variable regions of any chain and no two myeloma immunoglobulins or Bence Jones proteins have thus far been found to be identical in sequence (30). The constant region, however, is essentially no different from other proteins in that the variation in the amino acids found at any position is ascribable to species and class variations or to genetic variants such as Inv factors. By comparison of sequence data on the variable and constant regions of Bence Jones proteins with amino acid composition of purified human antibodies, it could be shown that most of the compositional variation could only originate in the variable region (see Kabat in reference 18).

From sequence data, a variety of hypotheses have been advanced (7, 31–35) to explain the structural basis of antibody complementarity. All of these are selective theories, i.e. they consider that the information for complementarity is essentially built into the primary sequence of each chain and that a given antigen only triggers the biosynthesis of those antibody molecules having complementary receptor sites. There are two types of selective theories: germ line theories (36) and somatic mutation theories (37–39). At present no hypothesis is generally accepted. Excellent reviews (see above) are available.

The present communication is an extension of earlier efforts from this laboratory (18, p. 87, and 40–43) to locate more precisely those portions of the variable region which are directly responsible for antibody complementarity, that is which make direct contact with the antigenic determinant, and to explain the unique capacity of these proteins to have so many complementary regions.

As in the earlier studies, all human κ, human λ, and mouse κ Bence Jones protein and light chain sequences are aligned for maximum homology (44) and all variable regions are considered as a unit and compared with the constant regions. These earlier studies had called attention to the following:

(a) The variable regions had few if any species-specific positions while the constant regions of the human and mouse proteins had 36 species-specific amino acid substitutions per 107 residues (40, 45). A species-specific position is defined as one at which the amino acid residues in the mouse proteins differ from those in the human proteins.

(b) When the invariant residues of these two regions were compared, the latest tabulation (45) showed the variable regions to have 10 invariant and almost invariant glycines and no invariant alanines, leucines, valines, histi-

dines, lysines, or serines while the constant regions had 3 each of invariant
alanine, leucine, and valine, and 2 invariant histidines, 2 invariant lysines, and
5 invariant serines. It was suggested that the invariant glycines were important
in contributing to the flexibility needed by the variable region in accommo-
dating the numerous substitutions (41, 43) at the variable positions. It was
also suggested that the invariant glycines near the end of the variable region
at positions 99 and 101, plus the almost invariant glycine at position 100,
provided a pivot upon which the complementarity-determining regions might
move to make better contact with the antigenic determinant (43; 18, p. 87)
just as the walls of the lysozyme site have been shown to adjust somewhat to
accommodate its hexasaccharide substrate (46). The hydrophobic residues in
the constant region were hypothesized to be involved in noncovalent bonding
to the heavy chain.

(c) From an examination of sequences of the κI, κII, and κIII subgroups
(Hood et al. in reference 16) (47, 48) of the human Bence Jones proteins in
which many of the proteins in a subgroup had an identical sequence for the
first 20–24 residues, it was postulated that there are two kinds of residues in
the variable regions, those making direct contact with the antigenic determi-
nant (complementarity determining) and those which are involved only in
three-dimensional folding (42). The latter would be expected to have less
stringent requirements, and more mutation noise would be permitted than
with the complementarity-determining residues. This distinction led to the
inspection of the sequences for short stretches showing very high variability
and two of these were identified: the most variable beginning at residue 89
and ending at 97, the other running from residue 24 through 34. Each of these
two unusually highly variable regions began after an invariant half-cystine
and was followed by an invariant phenylalanine (residue 98) and an invariant
tryptophane (residue 35) respectively. It is of interest that the two regions are
brought close together by the S—S bond I_{23}–II_{88} (45). Milstein (47), Milstein
and Pink (7), and Franĕk (49) have also called attention to the highly variable
positions in these regions and Franĕk (49) has noted an additional highly
variable region around residues 52–55. It was hypothesized (45) that these first
two regions might represent the complementarity-determining regions and
that complementarity might be acquired by the insertion of small linear se-
quences into the light and heavy chains by some episomal or other insertion
mechanism. It is striking that the differences in chain length seen in the Bence
Jones proteins are confined to these two regions of the chain. The remaining
portions of each chain would be essentially under the control of structural
genes. The inserted sequences would be drawn from a large but finite set and
either inserted under the influence of antigen, if antibody-forming cells are
multipotent, or individual sequences might be distributed to immunoglobulin-

forming cells during differentiation if the capacity of individual cells to synthesize antibody is restricted.

This working hypothesis offers several advantages:

(a) It is capable of providing the evolutionary stability and accounts for the universality of the antibody-forming mechanism throughout the vertebrates. Germ line theories (34–36) postulate one gene for each of the thousand or more variable regions (30). This would be expected to result in divergence during evolution since the loss by mutation of any one variable region would only minimally affect the capacity to form antibody and survival; thus individuals and populations lacking certain variable regions would arise.

(b) It offers a substantial simplification to the problem of producing a very large number of complementary sites. While it is known that in all proteins with specific receptors the site is formed by residues from widely separated portions of the chain, these sites are all formed by single chains. Thus, forming a three-dimensional site must involve residues from various regions. The antibody site being formed by a heavy and a light chain need not necessarily be so restricted.

Since much additional data on the light chains and a number of heavy chain sequences have been accumulated, the present communication represents a further attempt at analyzing the unique features of the variable regions of immunoglobulin chains. Among aspects considered are the role of glycine, invariant residues, and hydrophobicity patterns, and the highly variable portions, with a view to localizing the regions responsible for complementarity and evaluating various theories in terms of evolutionary origin and perpetuation of the antibody-forming mechanism.

Sequence Data Employed—Complete and partial sequence data have been published on 77 Bence Jones proteins and immunoglobulin light chains as well as on a number of heavy chains. Data were available on 24 human κI, 4 human κII, 17 human κIII, 10 human λI, 2 human λII, 6 human λIII, 5 human λIV, 2 human λV, 2 mouse κI, and 5 mouse κII proteins.[1]

The original light chain sequence data may be found in the following references.

HBJ 98: Baglioni, C. 1967. *Biochem. Biophys. Res. Commun.* **26**:82.

Eu: Cunningham, B. A., P. D. Gottlieb, W. H. Konigsberg, and G. M. Edelman. 1968. *Biochemistry.* **7**:1983.

Mil (human κII): Dreyer, W. J., W. R. Gray, and L. Hood. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:353.

Hac, Dob, Pal: Grant, A., and L. Hood. Unpublished work.

Roy, Cum: Hilschman, N., and L. C. Craig. 1965. *Proc. Nat Acad. Sci. U. S. A.* **53**:1403; Hilschmann, N. 1967. *Hoppe-Seyler's Z. Physiol. Chem.* **348**:1077; Hilschmann, N., H. U. Barnikol, M. Hess, B. Langer, H. Ponstingl, M. Steinmetz-Kayne, L. Suter, and S. Watanabe. 1968. *Fed. Eur. Biochem. Soc. Symp., 5th.* In press.

---

[1] The World Health Organization has recently changed the notation of subgroups so that human κII in this paper will become human κIII and human κIII will become human κII.

HS 78, HS 92, HS 94, HS 68, HS 70, HS 77, HS 86, HS 24: Hood, L., and D. Ein. 1968. *Nature (London)*. **220**:764.

HBJ 7, HBJ 11, HBJ 2, HBJ 8: Hood, L., W. R. Gray, and W. J. Dreyer. 1966. *J. Mol. Biol.* **22**:179.

MBJ 41, MBJ 70, MBJ 6: Hood, L., W. R. Gray, and W. J. Dreyer. 1966. *Proc. Nat'l Acad. Sci. U. S. A.* **55**:826.

HBJ 10, HBJ 1, HBJ 4, HBJ 6, HBJ 5, HS 4, HBJ 12, HS 6, HBJ 15: Hood, L., W. R. Gray, B. G. Sanders, and W. J. Dreyer. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:133.

Ste: Edman, P., and A. G. Cooper. 1968 *Fed. Eur. Biochem Soc. Letters*. **2**:33; Hood, L., and D. W. Talmage. 1969. *In* Developmental Aspects of Antibody Formation and Structure. Prague. In press.

Lay, Mar, Ioc, Wag, How, Koh: Kaplan, A. P. and H. Metzger. 1969. *Biochemistry*. **10**:3944.

New, III, Mil (human λIV): Langer, B., M. Steinmetz-Kayne, and N. Hilschmann. 1968. *Hoppe-Seyler's Z. Physiol. Chem.* **349**:945.

BJ, Ker: Milstein, C. 1966. *Biochem. J.* **101**:352.

Rad, Fr4: Milstein, C. 1967. *Nature (London)* **216**:330.

X: Milstein, C. 1968. *Biochem. J.* **110**:631.

Bel, Man, B6: Milstein, C. 1968. *Fed. Eur. Biochem. Soc. Symp. on γ-globulin*, Prague.

Day, MBJ46, Roy: Atlas of Protein Sequence and Structure, M. O. Dayhoff, Editor. 1969.

Mz: Milstein, C., B. Frangione, and J. R. L. Pink. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:31.

Ale, Car, Dee: Milstein, C., C. P. Milstein, and A. Feinstein. 1969. *Nature (London)* **221**:151.

Cra, Pap, Lux, Mon, Con, Tra, Nig, Win, Gra, Cas, Smi: Niall, H., and P. Edman. 1967. *Nature (London)* **216**:262.

MOPC 149, AdjPC 9, MOPC 157: Perham, R., E. Appella, and M. Potter. 1966. *Science (Washington)* **154**:391.

Kern: Ponstingl, H., M. Hess, and N. Hilschmann. 1968. *Hoppe-Seyler's Z. Physiol. Chem.* **349**:867.

Tew: Putnam, F. W. 1969. *Science (Washington)*. **163**:633.

Ag, Ha, Bo, Sh: Putnam, F. W., K. Titani, M. Wikler, and T. Shinoda. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:9; Titani, K., T. Shinoda, and F. W. Putnam. 1969. *J. Biol. Chem.* **244**:3550.

TI: Suter, L., H. U. Barnikol, S. Watanabe, and N. Hilschmann. 1969. *Hoppe-Seyler's Z. Physiol. Chem.* **350**:275.

The accumulation of such large numbers of sequences makes it possible to use statistical criteria in defining the types of residues. Thus in earlier studies, an invariant residue was rigidly defined, e.g., a position at which all samples showed the same amino acid residue sometimes allowing a single exception. The definition of an invariant residue used in this paper is taken as a position at which 88–90% or more of the samples contain the same amino acid. This may allow potential functions to be recognized despite possible errors or uncertainties in sequence, or occasional substitutions compatible with function.

A summary of the sequence data is provided in Table I which lists the amino acids found at any position in any subgroup of human κ-, human λ-, and mouse κ-chains, the number of times each occurs, and the total number of sequences

TABLE I

Amino Acids Found at each Position in the Variable Region of the Various Subgroups of Human κ-, Human λ- and Mouse κ-Bence Jones Proteins

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
0	74	Glu	1	0	1	0	0	0	0	0	0	0	0
		---	73	23	3	17	9	2	6	4	2	2	5
1	74	Lys	1	0	0	1	0	0	0	0	0	0	0
		PCA	17	0	0	0	8	2	6	0	0	0	1
		Asp	29	18	4	1	0	0	0	0	0	2	4
		Asx	5	5	0	0	0	0	0	0	0	0	0
		Glu	15	0	0	14	1	0	0	0	0	0	0
		Glx	1	0	0	1	0	0	0	0	0	0	0
		---	6	0	0	0	0	0	0	4	2	0	0
2	73	Ile	48	22	3	16	0	0	0	0	0	2	5
		Tyr	4	0	0	0	0	0	0	4	0	0	0
		Val	1	1	0	0	0	0	0	0	0	0	0
		Met	1	0	0	1	0	0	0	0	0	0	0
		Ser	19	0	0	0	9	2	6	0	2	0	0
3	72	Ile	1	0	0	1	0	0	0	0	0	0	0
		Pro	1	0	0	0	0	0	1	0	0	0	0
		Leu	1	1	0	0	0	0	0	0	0	0	0
		Val	33	0	3	16	8	0	0	1	0	0	5
		Ala	9	0	0	0	1	2	5	1	0	0	0
		Asp	1	0	0	0	0	0	0	1	0	0	0
		Glu	2	0	0	0	0	0	0	1	1	0	0
		Gln	21	19	0	0	0	0	0	0	0	2	0
		Glx	3	3	0	0	0	0	0	0	0	0	0
4	71	Leu	43	3	1	14	9	2	6	4	1	0	3
		Val	2	0	0	1	0	0	0	0	0	0	1
		Met	26	20	2	2	0	0	0	0	0	2	0
5	70	Ala	2	0	0	0	0	0	2	0	0	0	0
		Thr	67	23	3	17	9	2	3	4	1	2	3
		Ser	1	0	0	0	0	0	1	0	0	0	0
6	69	Gln	63	20	3	16	8	1	6	4	1	1	3
		Glx	6	3	0	1	1	1	0	0	0	0	0
7	63	Pro	20	0	0	0	9	2	6	3	0	0	0
		Thr	1	0	1	0	0	0	0	0	0	0	0
		Ser	41	22	2	14	0	0	0	0	0	1	2
		Asp	1	0	0	0	0	0	0	0	1	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
8	64	Pro	58	22	3	15	9	2	0	3	1	1	2
		Ala	6	0	0	0	0	0	6	0	0	0	0
9	63	Leu	3	0	3	0	0	0	0	0	0	0	0
		Ala	7	0	0	4	0	0	0	0	1	0	2
		Thr	1	1	0	0	0	0	0	0	0	0	0
		Ser	41	21	0	0	9	2	6	2	0	1	0
		Gly	10	0	0	10	0	0	0	0	0	0	0
		Asx	1	0	0	1	0	0	0	0	0	0	0
10	63	Phe	1	1	0	0	0	0	0	0	0	0	0
		Thr	17	3	0	14	0	0	0	0	0	0	0
		Ser	25	18	4	0	0	0	0	0	0	1	2
		---	20	0	0	0	9	2	6	2	1	0	0
11	63	Leu	43	22	4	14	0	0	0	0	0	1	2
		Val	15	0	0	0	6	0	6	2	1	0	0
		Ala	5	0	0	0	3	2	0	0	0	0	0
12	61	Pro	4	0	4	0	0	0	0	0	0	0	0
		Ala	1	0	0	0	0	0	0	0	0	0	1
		Ser	56	21	0	13	9	2	6	2	1	1	1
13	61	Leu	12	0	0	12	0	0	0	0	0	0	0
		Val	11	2	4	0	0	0	0	2	1	0	2
		Met	1	0	0	1	0	0	0	0	0	0	0
		Ala	23	19	0	0	3	0	0	0	0	1	0
		Gly	14	0	0	0	6	2	6	0	0	0	0
14	61	Ala	6	0	0	0	5	0	0	0	1	0	0
		Thr	9	0	4	0	4	0	0	0	0	0	1
		Ser	46	21	0	13	0	2	6	2	0	1	1
15	61	Pro	36	0	4	13	8	2	6	2	0	0	1
		Leu	4	1	0	0	0	0	0	0	1	1	1
		Val	20	20	0	0	0	0	0	0	0	0	0
		Asx	1	0	0	0	1	0	0	0	0	0	0
16	61	Arg	1	1	0	0	0	0	0	0	0	0	0
		Gly	60	20	4	13	9	2	6	2	1	1	2
17	61	Asp	23	21	0	1	0	0	0	0	0	0	1
		Glu	17	0	4	11	0	0	1	0	0	1	0
		Gln	19	0	0	0	8	2	5	2	1	0	1
		Glx	2	0	0	1	1	0	0	0	0	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studies	Amino Acids	Total	Human Kappa I	II	III	Human Lambda I	II	III	IV	V	Mouse I	Kappa II
18	61	Pro	4	0	4	0	0	0	0	0	0	0	0
		Lys	1	0	0	0	1	0	0	0	0	0	0
		Ala	1	0	0	0	1	0	0	0	0	0	0
		Arg	40	21	0	13	4	0	0	0	0	1	1
		Thr	5	0	0	0	1	0	1	2	1	0	0
		Ser	9	0	0	0	1	2	5	0	0	0	1
		Gly	1	0	0	0	1	0	0	0	0	0	0
19	53	Ile	7	1	0	0	0	0	6	0	0	0	0
		Val	28	17	0	0	6	2	0	0	1	1	1
		Ala	18	0	4	10	1	0	0	2	0	0	1
20	53	Ile	1	0	0	0	1	0	0	0	0	0	0
		Val	1	0	0	0	0	0	0	1	0	0	0
		Ala	3	1	0	2	0	0	0	0	0	0	0
		Arg	1	0	0	0	0	0	0	0	1	0	0
		Thr	40	17	0	8	6	2	6	0	0	0	1
		Ser	7	0	4	0	0	0	0	1	0	1	1
21	43	Ile	30	14	4	0	5	2	0	3	1	0	1
		Leu	12	1	0	9	0	0	0	0	0	1	1
		Val	1	0	0	1	0	0	0	0	0	0	0
22	42	Ala	1	1	0	0	0	0	0	0	0	0	0
		Thr	19	14	0	0	0	0	0	3	1	1	0
		Ser	22	0	4	9	5	2	0	0	0	0	2
23	30	Cys	30	9	3	4	5	2	0	3	1	1	2
24	26	Arg	11	1	3	4	0	0	0	0	0	1	2
		Thr	2	0	0	0	0	2	0	0	0	0	0
		Ser	6	0	0	0	4	0	0	2	0	0	0
		Gly	1	0	0	0	0	0	0	1	0	0	0
		Gln	4	3	0	0	0	0	0	0	1	0	0
		Glx	2	2	0	0	0	0	0	0	0	0	0
25	25	Ala	13	6	0	4	0	0	0	0	0	1	2
		Ser	2	0	2	0	0	0	0	0	0	0	0
		Gly	10	0	0	0	4	2	0	3	1	0	0
26	24	Thr	2	0	0	0	0	2	0	0	0	0	0
		Ser	16	6	2	4	1	0	0	0	0	1	2
		Gly	2	0	0	0	2	0	0	0	0	0	0
		Asp	4	0	0	0	0	0	0	3	1	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
27	23	Lys	1	0	0	0	0	0	0	1	0	0	0
		Ser	5	0	0	0	3	1	0	0	1	0	0
		Asn	1	0	0	0	0	0	0	1	0	0	0
		Glu	1	0	0	0	0	0	0	1	0	0	0
		Gln	13	5	1	4	0	0	0	0	0	1	2
		Glx	2	1	1	0	0	0	0	0	0	0	0
a	22	Ser	5	0	1	3	0	0	0	0	0	0	1
		Asn	1	0	1	0	0	0	0	0	0	0	0
		---	16	6	0	1	3	1	0	2	1	1	1
b	22	Leu	2	0	2	0	0	0	0	0	0	0	0
		Val	1	0	0	0	0	0	0	0	0	0	1
		---	19	6	0	4	3	1	0	2	1	1	1
c	22	Leu	2	0	2	0	0	0	0	0	0	0	0
		---	20	6	0	4	3	1	0	2	1	1	2
d	22	Thr	1	0	0	0	1	0	0	0	0	0	0
		Ser	3	0	0	0	2	1	0	0	0	0	0
		Asp	1	0	1	0	0	0	0	0	0	0	0
		---	17	6	1	4	0	0	0	2	1	1	2
e	22	Asp	2	0	1	0	0	1	0	0	0	0	0
		Asn	3	0	0	0	3	0	0	0	0	0	0
		Asx	1	0	0	0	0	0	0	0	0	0	1
		---	16	6	1	4	0	0	0	2	1	1	1
f	22	Val	2	0	0	1	0	1	0	0	0	0	0
		Ser	2	0	1	0	0	0	0	0	0	0	1
		Gly	1	0	0	0	1	0	0	0	0	0	0
		Glx	1	0	1	0	0	0	0	0	0	0	0
		---	16	6	0	3	2	0	0	2	1	1	1
28	22	Ile	1	0	0	0	1	0	0	0	0	0	0
		Leu	4	0	0	1	0	0	0	2	1	0	0
		Val	2	0	0	2	0	0	0	0	0	0	0
		Met	1	0	0	0	1	0	0	0	0	0	0
		Thr	1	0	0	0	1	0	0	0	0	0	0
		Ser	3	1	1	1	0	0	0	0	0	0	0
		Gly	2	0	1	0	0	1	0	0	0	0	0
		His	1	0	0	0	0	0	0	0	0	0	1
		Asp	4	4	0	0	0	0	0	0	0	0	0
		Asx	3	1	0	0	0	0	0	0	0	1	1

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa I	II	III	Human Lambda I	II	III	IV	V	Mouse Kappa I	II
29	21	Ile	8	6	0	0	0	0	0	0	0	1	1
		Arg	2	0	0	1	0	0	0	0	1	0	0
		Ser	3	0	0	3	0	0	0	0	0	0	0
		Gly	4	0	0	0	2	0	0	1	0	0	1
		Glu	1	0	0	0	0	0	0	1	0	0	0
		Asp	1	0	1	0	0	0	0	.0	0	0	0
		Asx	2	0	1	0	0	1	0	0	0	0	0
30	21	Ile	2	1	0	0	0	0	0	0	0	0	1
		Lys	2	1	0	0	0	0	0	1	0	0	0
		Ser	2	1	0	0	0	0	0	0	0	0	1
		Gly	5	0	2	1	0	0	0	0	1	1	0
		Asp	1	0	0	0	0	0	0	1	0	0	0
		Asn	8	3	0	3	2	0	0	0	0	0	0
		Asx	1	0	0	0	0	1	0	0	0	0	0
31	20	Tyr	1	0	0	0	0	0	0	0	1	0	0
		Phe	1	0	0	0	0	0	0	1	0	0	0
		Lys	2	1	0	0	0	1	0	0	0	0	0
		Thr	2	1	1	0	0	0	0	0	0	0	0
		Ser	7	1	0	2	2	0	0	0	0	1	1
		His	1	1	0	0	0	0	0	0	0	0	0
		Asn	4	1	0	2	0	0	0	0	0	0	1
		Asx	2	1	1	0	0	0	0	0	0	0	0
32	20	Trp	1	1	0	0	0	0	0	0	0	0	0
		Tyr	10	2	2	3	2	1	0	0	0	0	0
		Phe	5	3	0	1	0	0	0	0	0	0	1
		Leu	1	0	0	0	0	0	0	0	0	1	0
		Thr	1	0	0	0	0	0	0	1	0	0	0
		Asp	1	0	0	0	0	0	0	0	1	0	0
		Asn	1	0	0	0	0	0	0	0	0	0	1
33	18	Leu	11	4	2	4	0	0	0	0	0	0	1
		Val	4	0	0	0	2	1	0	1	0	0	0
		Met	1	0	0	0	0	0	0	0	0	0	1
		Ala	1	0	0	0	0	0	0	0	1	0	0
		Ser	1	0	0	0	0	0	0	0	0	1	0
34	18	Tyr	1	0	0	0	1	0	0	0	0	0	0
		Ala	6	1	0	4	0	0	0	0	1	0	0
		Ser	3	0	0	0	1	1	0	1	0	0	0
		Asp	1	0	1	0	0	0	0	0	0	0	0
		Asn	4	2	1	0	0	0	0	0	0	0	1
		Asx	3	1	0	0	0	0	0	0	0	1	1

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
35	17	Trp	17	4	2	4	2	1	0	1	1	1	1
36	17	Tyr	13	4	2	4	1	1	0	0	1	0	0
		Phe	2	0	0	0	0	0	0	1	0	0	1
		Leu	1	0	0	0	0	0	0	0	0	1	0
		His	1	0	0	0	1	0	0	0	0	0	0
37	16	Leu	2	0	2	0	0	0	0	0	0	0	0
		Gln	12	3	0	4	2	1	0	1	1	0	0
		Glx	2	0	0	0	0	0	0	0	0	1	1
38	16	His	1	0	0	0	1	0	0	0	0	0	0
		Gln	12	3	1	4	1	1	0	1	1	0	0
		Glx	3	0	1	0	0	0	0	0	0	1	1
39	16	Leu	2	0	0	0	2	0	0	0	0	0	0
		Lys	9	2	2	3	0	0	0	0	1	0	1
		Arg	2	0	0	1	0	0	0	1	0	0	0
		Gly	1	1	0	0	0	0	0	0	0	0	0
		His	1	0	0	0	0	1	0	0	0	0	0
		Asx	1	0	0	0	0	0	0	0	0	1	0
40	16	Pro	15	3	1	4	2	1	0	1	1	1	1
		Ala	1	0	1	0	0	0	0	0	0	0	0
41	16	Lys	1	1	0	0	0	0	0	0	0	0	0
		Gly	15	2	2	4	2	1	0	1	1	1	1
42	16	Lys	3	3	0	0	0	0	0	0	0	0	0
		Arg	1	0	0	0	0	1	0	0	0	0	0
		Thr	2	0	0	0	2	0	0	0	0	0	0
		Gln	7	0	1	4	0	0	0	1	1	0	0
		Glx	3	0	1	0	0	0	0	0	0	1	1
43	16	Pro	1	0	0	0	0	0	0	0	0	0	1
		Ala	11	3	0	4	2	1	0	0	1	0	0
		Thr	1	0	0	0	0	0	0	0	0	1	0
		Ser	2	0	1	0	0	0	0	1	0	0	0
		Gln	1	0	1	0	0	0	0	0	0	0	0
44	16	Ile	1	0	0	0	0	0	0	0	0	1	0
		Pro	15	3	2	4	2	1	0	1	1	0	1
45	16	Leu	2	0	0	0	0	0	0	1	1	0	0
		Lys	9	3	0	1	2	1	0	0	0	1	1
		Arg	3	0	0	3	0	0	0	0	0	0	0
		Ser	1	0	1	0	0	0	0	0	0	0	0
		Glx	1	0	1	0	0	0	0	0	0	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
46	14	Ile	1	1	0	0	0	0	0	0	0	0	0
		Leu	12	2	2	2	2	1	0	1	1	0	1
		Arg	1	0	0	0	0	0	0	0	0	1	0
47	14	Leu	11	3	2	2	2	0	0	0	0	1	1
		Val	3	0	0	0	0	1	0	1	1	0	0
48	14	Ile	12	2	2	1	2	1	0	1	1	1	1
		Met	2	1	0	1	0	0	0	0	0	0	0
49	14	Tyr	13	3	2	2	2	0	0	1	1	1	1
		Phe	1	0	0	0	0	1	0	0	0	0	0
50	14	Leu	1	0	1	0	0	0	0	0	0	0	0
		Val	1	0	0	1	0	0	0	0	0	0	0
		Lys	1	1	0	0	0	0	0	0	0	0	0
		Ala	2	0	0	0	0	0	0	0	0	1	1
		Arg	1	0	0	0	1	0	0	0	0	0	0
		Thr	1	0	1	0	0	0	0	0	0	0	0
		Ser	1	0	0	0	0	0	0	1	0	0	0
		Gly	3	0	0	1	0	1	0	0	1	0	0
		Asp	2	2	0	0	0	0	0	0	0	0	0
		Glu	1	0	0	0	1	0	0	0	0	0	0
51	14	Leu	1	0	1	0	0	0	0	0	0	0	0
		Val	2	0	0	1	0	1	0	0	0	0	0
		Ala	5	3	0	1	0	0	0	0	0	0	1
		Arg	1	0	0	0	0	0	0	0	1	0	0
		Thr	1	0	0	0	0	0	0	0	0	1	0
		Gly	1	0	1	0	0	0	0	0	0	0	0
		His	1	0	0	0	0	0	0	1	0	0	0
		Asp	2	0	0	0	2	0	0	0	0	0	0
52	14	Ser	10	3	2	2	0	1	0	0	0	1	1
		Asp	1	0	0	0	1	0	0	0	0	0	0
		Asn	2	0	0	0	1	0	0	0	1	0	0
		Glu	1	0	0	0	0	0	0	1	0	0	0
53	14	Tyr	1	0	1	0	0	0	0	0	0	0	0
		Lys	3	1	0	0	2	0	0	0	0	0	0
		Ser	4	1	0	2	0	0	0	0	0	1	0
		Asn	4	1	1	0	0	0	0	0	1	0	1
		Glu	1	0	0	0	0	1	0	0	0	0	0
		Glx	1	0	0	0	0	0	0	1	0	0	0
54	14	Leu	4	3	0	0	0	0	0	0	0	1	0
		Arg	9	0	2	2	2	1	0	1	1	0	0
		Gln	1	0	0	0	0	0	0	0	0	0	1

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
55	16	Pro	5	0	0	0	2	1	0	1	1	0	0
		Ala	6	0	2	4	0	0	0	0	0	0	0
		Gly	1	0	0	0	0	0	0	0	0	0	1
		Asx	1	0	0	0	0	0	0	0	0	1	0
		Glu	3	3	0	0	0	0	0	0	0	0	0
56	16	Ala	1	1	0	0	0	0	0	0	0	0	0
		Thr	5	1	0	4	0	0	0	0	0	0	0
		Ser	10	1	2	0	2	1	0	1	1	1	1
57	16	Thr	1	0	0	0	0	0	0	1	0	0	0
		Gly	15	3	2	4	2	1	0	0	1	1	1
58	16	Ile	6	0	0	3	1	0	0	1	1	0	0
		Val	9	3	2	0	1	1	0	0	0	1	1
		Thr	1	0	0	1	0	0	0	0	0	0	0
59	16	Pro	16	3	2	4	2	1	0	1	1	1	1
60	16	Val	1	0	0	1	0	0	0	0	0	0	0
		Lys	1	0	0	0	0	0	0	0	0	1	0
		Ala	1	0	0	0	0	0	0	0	0	0	1
		Ser	3	3	0	0	0	0	0	0	0	0	0
		Asp	8	0	1	3	2	1	0	0	1	0	0
		Asn	1	0	1	0	0	0	0	0	0	0	0
		Glu	1	0	0	0	0	0	0	1	0	0	0
61	16	Arg	16	3	2	4	2	1	0	1	1	1	1
62	16	Ile	1	0	0	0	1	0	0	0	0	0	0
		Phe	15	3	2	4	1	1	0	1	1	1	1
63	16	Ile	1	1	0	0	0	0	0	0	0	0	0
		Ser	15	2	2	4	2	1	0	1	1	1	1
64	16	Ala	1	0	0	0	1	0	0	0	0	0	0
		Gly	15	3	2	4	1	1	0	1	1	1	1
65	16	Thr	1	1	0	0	0	0	0	0	0	0	0
		Ser	15	2	2	4	2	1	0	1	1	1	1
66	16	Lys	3	0	0	0	2	1	0	0	0	0	0
		Arg	1	0	0	0	0	0	0	0	0	1	0
		Ser	2	0	0	0	0	0	0	1	1	0	0
		Gly	10	3	2	4	0	0	0	0	0	0	1
67	16	Phe	1	1	0	0	0	0	0	0	0	0	0
		Ser	15	2	2	4	2	1	0	1	1	1	1

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
68	16	Gly	15	3	2	4	2	0	0	1	1	1	1
		Asn	1	0	0	0	0	1	0	0	0	0	0
69	16	Ala	2	0	0	1	0	0	0	1	0	0	0
		Thr	11	3	2	3	2	0	0	0	0	0	1
		Ser	1	0	0	0	0	0	0	0	0	1	0
		His	1	0	0	0	0	0	0	0	1	0	0
		Asp	1	0	0	0	0	1	0	0	0	0	0
70	16	Thr	3	0	0	0	0	1	0	1	1	0	0
		Ser	2	0	0	0	2	0	0	0	0	0	0
		Asp	9	2	1	4	0	0	0	0	0	1	1
		Asx	1	0	1	0	0	0	0	0	0	0	0
		Glu	1	1	0	0	0	0	0	0	0	0	0
71	16	Tyr	1	0	0	0	0	0	0	0	0	1	0
		Phe	10	3	2	4	0	0	0	0	0	0	1
		Ala	5	0	0	0	2	1	0	1	1	0	0
72	16	Thr	11	3	2	4	1	0	0	1	0	0	0
		Ser	5	0	0	0	1	1	0	0	1	1	1
73	16	Phe	2	2	0	0	0	0	0	0	0	0	0
		Leu	14	1	2	4	2	1	0	1	1	1	1
74	16	Lys	1	0	1	0	0	0	0	0	0	0	0
		Ala	1	0	0	0	1	0	0	0	0	0	0
		Thr	12	3	1	4	0	1	0	1	1	1	0
		Gly	1	0	0	0	1	0	0	0	0	0	0
		Asn	1	0	0	0	0	0	0	0	0	0	1
75	16	Ile	15	3	2	4	2	0	0	1	1	1	1
		Val	1	0	0	0	0	1	0	0	0	0	0
76	16	Thr	2	0	0	0	1	0	0	0	1	0	0
		Ser	13	3	2	4	1	1	0	1	0	1	0
		His	1	0	0	0	0	0	0	0	0	0	1
77	17	Pro	1	0	0	0	0	0	0	0	0	0	1
		Arg	6	0	2	4	0	0	0	0	0	0	0
		Ser	4	3	0	0	0	0	0	0	0	1	0
		Gly	6	1	0	0	2	1	0	1	1	0	0
78	17	Leu	13	4	1	4	2	1	0	0	0	1	0
		Val	1	0	1	0	0	0	0	0	0	0	0
		Met	1	0	0	0	0	0	0	0	0	0	1
		Ala	2	0	0	0	0	0	0	1	1	0	0

TAI TE WU AND ELVIN A. KABAT   225

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
79	17	Arg	3	0	0	0	2	1	0	0	0	0	0
		Glu	5	0	1	3	0	0	0	0	0	1	0
		Gln	6	4	0	0	0	0	0	1	1	0	0
		Glx	3	0	1	1	0	0	0	0	0	0	1
80	17	Pro	9	4	1	4	0	0	0	0	0	0	0
		Ala	3	0	1	0	0	1	0	0	1	0	0
		Thr	1	0	0	0	1	0	0	0	0	0	0
		Ser	3	0	0	0	1	0	0	1	0	1	0
		Glx	1	0	0	0	0	0	0	0	0	0	1
81	17	Val	1	0	0	0	0	0	0	1	0	0	0
		Ala	1	1	0	0	0	0	0	0	0	0	0
		Gly	1	0	0	0	1	0	0	0	0	0	0
		Asp	1	1	0	0	0	0	0	0	0	0	0
		Asx	1	0	0	0	0	0	0	0	0	0	1
		Glu	10	2	1	4	1	0	0	0	1	1	0
		Glx	2	0	1	0	0	1	0	0	0	0	0
82	17	Asp	14	4	1	4	2	0	0	1	1	1	0
		Asx	3	0	1	0	0	1	0	0	0	0	1
83	18	Ile	2	2	0	0	0	0	0	0	0	0	0
		Phe	8	2	1	4	0	0	0	0	0	1	0
		Val	1	0	1	0	0	0	0	0	0	0	0
		Thr	1	0	0	0	0	0	0	0	0	0	1
		Glu	5	0	0	0	2	0	0	2	1	0	0
		Glx	1	0	0	0	0	1	0	0	0	0	0
84	18	Val	1	0	0	0	0	0	0	0	0	1	0
		Ala	16	4	1	4	2	1	0	2	1	0	1
		Gly	1	0	1	0	0	0	0	0	0	0	0
85	18	Val	6	0	2	4	0	0	0	0	0	0	0
		Met	1	0	0	0	0	0	0	0	0	0	1
		Thr	4	4	0	0	0	0	0	0	0	0	0
		His	1	0	0	0	1	0	0	0	0	0	0
		Asp	5	0	0	0	1	0	0	2	1	1	0
		Asx	1	0	0	0	0	1	0	0	0	0	0
86	20	Tyr	20	6	2	4	2	1	0	2	1	1	1
87	20	Tyr	16	6	2	3	2	1	0	1	1	0	0
		Phe	3	0	0	1	0	0	0	1	0	0	1
		His	1	0	0	0	1	0	0	0	0	0	0
88	21	Cys	21	6	2	4	3	1	0	2	1	1	1

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
89	22	Leu	1	0	0	0	0	0	0	0	0	1	0
		Met	1	0	1	0	0	0	0	0	0	0	0
		Ala	2	0	0	0	2	0	0	0	0	0	0
		Ser	1	0	0	0	0	1	0	0	0	0	0
		Asn	1	0	0	0	0	0	0	0	1	0	0
		Gln	15	6	2	4	1	0	0	2	0	0	0
		Glx	1	0	0	0	0	0	0	0	0	0	1
90	22	Met	1	0	1	0	0	0	0	0	0	0	0
		Ala	3	0	0	0	2	0	0	1	0	0	0
		Thr	2	0	0	0	1	0	0	1	0	0	0
		Ser	2	0	0	0	0	1	0	0	1	0	0
		Gln	13	6	2	4	0	0	0	0	0	1	0
		Glx	1	0	0	0	0	0	0	0	0	0	1
91	22	Trp	5	0	0	0	3	0	0	2	0	0	0
		Tyr	12	5	1	4	0	1	0	0	0	1	0
		Phe	1	1	0	0	0	0	0	0	0	0	0
		Ala	1	0	1	0	0	0	0	0	0	0	0
		Arg	2	0	1	0	0	0	0	0	1	0	0
		Ser	1	0	0	0	0	0	0	0	0	0	1
92	21	Leu	2	0	2	0	0	0	0	0	0	0	0
		Val	1	0	0	0	0	1	0	0	0	0	0
		Lys	1	0	0	0	0	0	0	0	0	0	1
		Ala	1	0	0	0	0	0	0	0	0	1	0
		Gly	3	0	0	3	0	0	0	0	0	0	0
		Asp	9	4	0	0	2	0	0	2	1	0	0
		Asn	1	1	0	0	0	0	0	0	0	0	0
		Glu	3	1	1	1	0	0	0	0	0	0	0
93	21	Tyr	1	0	0	0	1	0	0	0	0	0	0
		Thr	4	1	1	1	0	0	0	1	0	0	0
		Ser	7	1	0	2	1	0	0	1	1	1	0
		Gly	1	0	0	1	0	0	0	0	0	0	0
		His	1	1	0	0	0	0	0	0	0	0	0
		Asp	1	1	0	0	0	0	0	0	0	0	0
		Asn	2	2	0	0	0	0	0	0	0	0	0
		Asx	1	0	0	0	0	1	0	0	0	0	0
		Glu	2	0	1	0	0	0	0	0	0	0	1
		Gln	1	0	1	0	0	0	0	0	0	0	0
94	21	Ile	2	0	1	0	0	0	0	1	0	0	0
		Leu	5	5	0	0	0	0	0	0	0	0	0
		Val	1	0	0	0	0	0	0	0	0	0	1
		Met	1	0	0	0	0	0	0	1	0	0	0
		Ala	1	0	1	0	0	0	0	0	0	0	0
		Arg	1	0	0	0	1	0	0	0	0	0	0
		Ser	8	0	1	4	1	0	0	0	1	1	0
		Asp	1	1	0	0	0	0	0	0	0	0	0
		Asx	1	0	0	0	0	1	0	0	0	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
95	21	Pro	14	5	3	4	0	0	0	0	0	1	1
		Leu	2	0	0	0	2	0	0	0	0	0	0
		Thr	1	0	0	0	0	0	0	1	0	0	0
		Ser	2	1	0	0	0	0	0	1	0	0	0
		Gly	1	0	0	0	0	0	0	0	1	0	0
		Asx	1	0	0	0	0	1	0	0	0	0	0
96	21	Trp	2	0	0	0	0	0	0	0	0	1	1
		Ile	1	0	1	0	0	0	0	0	0	0	0
		Tyr	2	1	1	0	0	0	0	0	0	0	0
		Phe	2	1	0	1	0	0	0	0	0	0	0
		Pro	1	1	0	0	0	0	0	0	0	0	0
		Leu	2	1	1	0	0	0	0	0	0	0	0
		Lys	2	1	0	0	0	0	0	0	1	0	0
		Arg	1	1	0	0	0	0	0	0	0	0	0
		Thr	1	0	0	1	0	0	0	0	0	0	0
		Ser	2	0	0	1	1	0	0	0	0	0	0
		Asn	1	0	0	0	1	0	0	0	0	0	0
		Asx	1	0	0	0	0	1	0	0	0	0	0
		Gln	1	0	0	1	0	0	0	0	0	0	0
		---	2	0	0	0	0	0	0	2	0	0	0
97	20	Phe	1	0	0	0	0	1	0	0	0	0	0
		Pro	1	0	0	1	0	0	0	0	0	0	0
		Met	1	1	0	0	0	0	0	0	0	0	0
		Ala	2	0	0	0	2	0	0	0	0	0	0
		Thr	12	4	3	3	0	0	0	0	0	1	1
		His	1	0	0	0	0	0	0	0	1	0	0
		---	2	0	0	0	0	0	0	2	0	0	0
a	19	Val	4	0	0	0	2	0	0	1	1	0	0
		Ala	1	0	0	0	0	0	0	1	0	0	0
		---	14	5	3	4	0	0	0	0	0	1	1
b	20	Ile	1	0	0	0	0	0	0	1	0	0	0
		Leu	1	0	0	0	0	0	0	0	1	0	0
		Val	4	0	0	0	2	1	0	1	0	0	0
		---	14	5	3	4	0	0	0	0	0	1	1
98	20	Phe	20	5	3	4	2	1	0	2	1	1	1
99	20	Gly	20	5	3	4	2	1	0	2	1	1	1
100	20	Pro	1	1	0	0	0	0	0	0	0	0	0
		Gly	11	2	1	0	2	1	0	2	1	1	1
		Gln	8	2	2	4	0	0	0	0	0	0	0
101	19	Gly	19	4	3	4	2	1	0	2	1	1	1

TABLE I—*Concluded*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
102	19	Thr	18	4	3	3	2	1	0	2	1	1	1
		Ser	1	0	0	1	0	0	0	0	0	0	0
103	19	Lys	14	4	1	3	1	1	0	1	1	1	1
		Arg	3	0	1	1	0	0	0	1	0	0	0
		Asn	1	0	1	0	0	0	0	0	0	0	0
		Gln	1	0	0	0	1	0	0	0	0	0	0
104	22	Leu	15	3	2	3	1	1	0	2	1	1	1
		Val	7	3	2	1	1	0	0	0	0	0	0
105	22	Thr	6	0	0	0	2	1	0	2	1	0	0
		Asp	4	3	0	1	0	0	0	0	0	0	0
		Glu	12	3	4	3	0	0	0	0	0	1	1
106	22	Ile	12	3	4	3	0	0	0	0	0	1	1
		Phe	1	1	0	0	0	0	0	0	0	0	0
		Leu	2	1	0	1	0	0	0	0	0	0	0
		Val	7	1	0	0	2	1	0	2	1	0	0
a	22	Leu	6	0	0	0	2	1	0	2	1	0	0
		---	16	6	4	4	0	0	0	0	0	1	1
107	22	Lys	15	6	3	4	0	0	0	0	0	1	1
		Arg	3	0	1	0	1	1	0	0	0	0	0
		Ser	2	0	0	0	0	0	0	2	0	0	0
		Gly	2	0	0	0	1	0	0	0	1	0	0

studied at the given position. Only data for which the sequence has been clearly assigned by the various authors have been included.

The Role of Glycine—It has been suggested that glycine plays a unique role in the structure of the variable region of immunoglobulin light chains (18, p. 87; 41–43, 45). Jukes (50) and Welscher (51) have generally agreed with this. A further careful analysis becomes essential for the understanding of the function of the glycines in the over-all structure and in relation to antibody-combining sites.

The basic property that differentiates glycine from all other amino acids structurally is the absence of a side chain. As a result, glycine can have many sterically allowable configurations. This has been verified experimentally in the case of lysozyme (46, 52) and tosyl-α-chymotrypsin (53). The two angles, Φ and ψ (54), which specify the conformation of the backbone of an amino acid have been calculated for each of the amino acids from the known tertiary structures of lysozyme (46, 52) and of tosyl-α-chymotrypsin (53). A typical plot of the permissible angles of the glycine as compared with the alanine residues is shown in Fig. 1. The allowable configurations of alanine are mostly
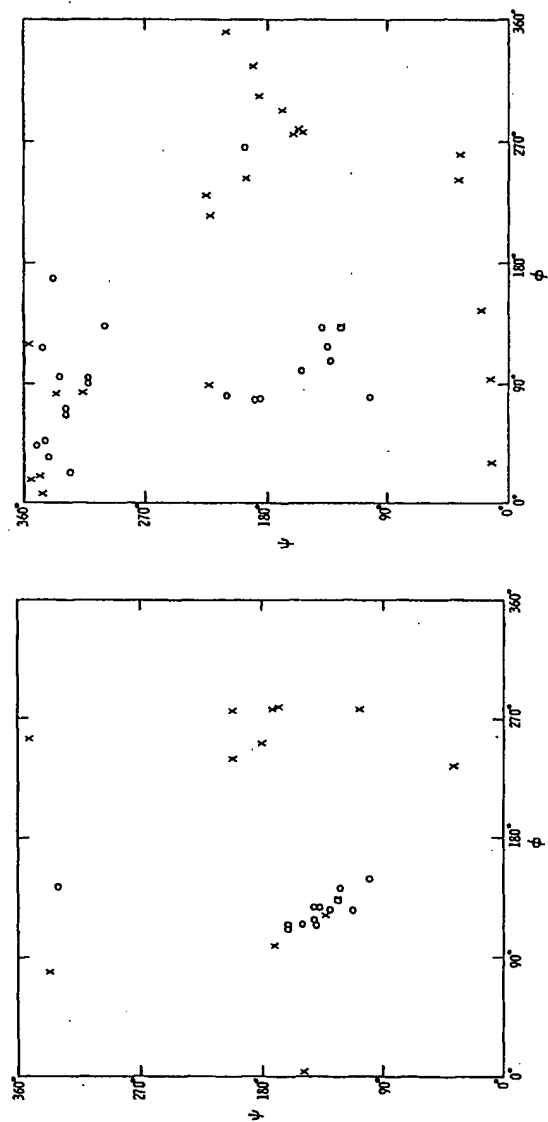
FIG. 1 b

FIG. 1 a

FIG. 1. Comparison of ϕ and ψ for glycine ($\times$) and alanine ($\bigcirc$) based on their occurrence in (a) lysozyme (data from references 46, 52), and in (b) tosyl-α-chymotrypsin (calculated from data given in reference 53). α indicates the values of ϕ and ψ for the α-helix.

clustered near the α-helical region of lysozyme (Fig. 1 a and reference 55), while those of glycine are widely distributed. Comparison with similar maps for other amino acids in lysozyme also shows them to be more restricted. The data for tosyl-α-chymotrypsin also show that glycine may have many more conformations (Fig. 1 b). This unique property of glycine thus may permit relative motion of the chains attached to the two ends of the molecule. With immuno-

TABLE II

Frequencies of Glycine Residues at Various Positions in the Variable Region of Light Chains

Position	Human Kappa			Human Lambda					Mouse Kappa		Total	%
	I	II	III	I	II	III	IV	V	I	II		
9			10/15								10/63	16
13				6/9	2/2	6/6					14/61	23
16	20/21	4/4	13/13	9/9	2/2	6/6	2/2	1/1	1/1	2/2	60/61	99
24								1/3			1/26	4
25				4/4	2/2		3/3	1/1			10/25	40
26				2/3							2/24	8
27f				1/3							1/22	5
28		1/2			1/1						2/22	9
29				2/2		1/2				1/2	4/21	19
30		2/2	1/4					1/1	1/1		5/21	24
39	1/3										1/16	6
41	2/3	2/2	4/4	2/2	1/1		1/1	1/1	1/1	1/1	15/16	94
50			1/2		1/1		1/1				3/14	21
51		1/2									1/14	7
55										1/1	1/16	6
57	3/3	2/2	4/4	2/2	1/1			1/1	1/1	1/1	15/16	94
64	3/3	2/2	4/4	1/2	1/1		1/1	1/1	1/1	1/1	15/16	94
66	3/3	2/2	4/4							1/1	10/16	62
68	3/3	2/2	4/4	2/2			1/1	1/1	1/1	1/1	15/16	94
74				1/2							1/16	6
77	1/4			2/2	1/1		1/1	1/1			6/17	35
81				1/2							1/17	6
84		1/2									1/18	6

TABLE II—*Concluded*

Position	Human Kappa			Human Lambda					Mouse Kappa		Total	%
	I	II	III	I	II	III	IV	V	I	II		
92			3/4								3/21	14
93			1/4								1/21	5
95								1/1			1/21	5
99	5/5	3/3	4/4	2/2	1/1		2/2	1/1	1/1	1/1	20/20	100
100	2/5	1/3		2/2	1/1		2/2	1/1	1/1	1/1	11/20	55
101	4/4	3/3	4/4	2/2	1/1		2/2	1/1	1/1	1/1	19/19	100
107				1/2				1/1			2/18	11

Fractions represent the number of instances in which glycine occurs to the total number of proteins studied at the given position for each subgroup.
When values have not been given, glycine has not been reported at that position in the subgroup.

globulins, flexibility of the protein backbone can be one of the major factors that permit substitution of various amino acids at the variable positions; it also may allow movement of the site to make most favorable contact in combining with an antigenic determinant. Though a glycine residue confers maximum flexibility over all other amino acids, it might also arise from a random mutation in which the difference between glycine and other amino acids is not adverse for the over-all structure. In addition, some glycines might be complementarity determining. These latter two kinds of glycines must be distinguished from the first.

For the variable region of the light chains of human and mouse immunoglobulins and of Bence Jones proteins, alignment of amino acid sequences serves to identify the glycines which may be conferring flexibility as shown in Table II. All the glycines are listed. The frequencies, expressed as per cent, can roughly be divided into three categories:

A. 94-100%: Positions 16, 41 (or 39), 57, 64, 68, 99, and 101. Since glycine occurs at these positions in nearly all the proteins studied, it must have a fundamental structural significance and has been preserved in the evolution to man and mouse. These glycines are assumed to confer flexibility unique to antibodies.

It is of interest that glycine occurs at position 41 in 15 of 16 proteins studied (94%). The sequence at residues 39, 40, and 41 is Lys-Pro-Gly in 14 cases and Lys-Ala-Gly in one case. In the single exception, a human κI protein Ag, the

sequence at 39, 40, and 41 is Gly-Pro-Lys.[1] Thus if the reported sequence is correct, the glycine at residue 39 might well serve the same function as the glycine at position 41. The two positions 41 and 39 may thus provide one invariant glycine (100%).

B. 35–62%: Positions 25, 66, 77, and 100. A careful examination indicates that glycines at these positions are at least group specific. Thus, within a group (κ or λ), they could serve the same function as the glycines of category *A*.

C. 4–24%: Positions 9, 13, 24, 26, 27f, 28, 29, 30, 50, 51, 55, 74, 81, 84, 92, 93, 95, and 107. These glycine residues are at variable positions. They therefore play a distinctly different role from those of the other two categories. They might either be related to antibody complementarity or, if not involved in the site itself, could have arisen from random mutation and be nevertheless compatible with three-dimensional folding.

Thus there are about 8 (human κ and mouse κ) to 10 (human λ) glycines in the variable region of the light chains of all proteins for which sequence data are available at these positions. They are as follows:

Human κ: Positions 16,   41 (or 39), 57, 64, 66, 68,   99,   and 101.
Human λ: Positions 16, 25, 41,   57, 64,   68, 77, 99, 100, and 101.
Mouse κ: Positions 16,   41,   57, 64,   68,   99, 100, and 101.

Positions 99, 100, and 101 have been postulated to function as a pivot permitting the combining regions of the light and heavy chains to make most favorable contact with the antigenic determinant (18, p. 87; and 41, 43). Examination of the heavy chain sequences reported to date (56, 57) shows Gly-Gln-Gly at positions 112, 113, and 114 in two instances (He and Daw), Gly-Arg-Gly in one (Cor) and Gly-Gly at positions 114 and 115 in Eu; but in the latter protein two gaps have been placed at positions 108 and 109 in aligning with He for maximum homology. Thus these glycines could also be functionally and positionally equivalent.

Invariant Residues—Earlier comparisons of the invariant residues of the variable and constant regions were based entirely on those positions at which only a single amino acid occurred. As more data accumulated this number diminished, until there are now only 11 such positions in the variable region: Gln 6, Cys 23, Trp 35, Pro 59, Arg 61, Asp (Asx) 82, Tyr 86, Cys 88, Phe 98, Gly 99, and Gly 101. However, if one accepts as essentially invariant those positions at which more than 88–90% of the proteins studied have the same amino acid at a given position, this number increases to 29. A comparison of these residues with those in the constant region is given in Table III. This procedure allows for possible errors as well as for the ability of some residues to substitute for others at a given position. The difference between the con-

---

[1] The sequence of Roy was originally reported as Gly-Pro-Lys but has been changed to Lys-Pro-Gly (see references to Roy in sequence data).

stant and variable regions is not much different than originally appeared (41) except that there is one invariant alanine and one invariant leucine in the variable region. However, the difference between the two regions is still quite clear, the variable region having in addition no invariant valine, lysine, and

TABLE III

Comparison of the Invariant Residues of the Variable with those of the Constant Region in Human κ-, Human λ-, and Mouse κ-Bence Jones Proteins

Amino Acid	Variable Region	Constant Region
Gly	7*	0
Ala	1	3
Leu	1	3
Val	0	3
Lys	0	2
His	0	2
Ile	1	0
Ser	3	5
Glu	0	2
Gln	2	0
Arg	1	0
Pro	3	4
Tyr	2	3
Cys	2	3
Phe	2	2
Trp	1	1
Thr	2	1
Asp	1	1
Total	29	35

* Not including positions 25, 66, 77, and 100.

Invariant residues are those in which 88–90% of the proteins analyzed contain the same amino acid residue at a given position.

histidine, while the invariant residues in the constant region include three alanines, three leucines, three valines, two lysines, and two histidines.

Hydrophobicity Distribution of the Invariant Residues of the Variable Region— A parameter, $H\emptyset_{ave}$, based on the free energies of transfer of amino acid side chains from an organic to an aqueous environment has been introduced by Tanford (58) and applied by Bigelow (59) to the study of various proteins. $H\emptyset_{ave}$ is expressed in kilocalories per residue and varies from 3.00 for Trp to 0.45 for Thr and is very small, zero, or negative for Gly, Ser, His, Asp, Glu, Asn, and Gln; these have been taken as zero. These values have been used in

an examination of the invariant residues of the variable region. Table IV summarizes the findings and tabulates $H\emptyset_{ave}$ for the invariant residues and for the occasional other substituents found. In addition we have included data on two positions, 46 and 48, in which another substituent occurred with slightly lower frequency than that required by the definition of an invariant residue, and on three other positions 8, 12, and 37 in which the other substituents were confined to a single subgroup.

TABLE IV

Hydrophobicity, $H\emptyset_{ave}$, of the Invariant Residues and almost Invariant Residues of the Variable Region

| Position | Amino Acids | | Hydrophobicity, $H\emptyset_{ave}$ | |
	Invariant	Other	Invariant	Other
5	Thr 67*	Ala 2, Ser 1	0.45	0.75, 0.00
6	Gln 63		0.00	
16	Gly 60	Arg 1	0.00	0.75
23	Cys 30		1.00	
35	Trp 17		3.00	
38	Gln 12, Glx 3	His 1	0.00	0.00
40	Pro 15	Ala 1	2.60	0.75
41(39)	Gly 15	Lys 1	0.00	1.50
44	Pro 15	Ile 1	2.60	2.95
49	Tyr 13	Phe 1	2.85	2.65
57	Gly 15	Thr 1	0.00	0.45
59	Pro 16		2.60	
61	Arg 16		0.75	
62	Phe 15	Ile 1	2.65	2.95
63	Ser 15	Ile 1	0.00	2.95
64	Gly 15	Ala 1	0.00	0.75
65	Ser 15	Thr 1	0.00	0.45
67	Ser 15	Phe 1	0 00	2.65
68	Gly 15	Asn 1	0.00	0.00
73	Leu 14	Phe 2	2.40	2.65
75	Ile 15	Val 1	2.95	1.70
82	Asp 14, Asx 3		0.00	
84	Ala 16	Gly 1, Val 1	0.75	0.00, 1.70
86	Tyr 20		2.85	
88	Cys 21		1.00	
98	Phe 20		2.65	
99	Gly 20		0.00	
101	Gly 19		0.00	
102	Thr 18	Ser 1	0.45	0.00

TABLE IV—*Concluded*

Borderline for almost invariant.

Position	Amino Acids		Hydrophobicity, $H\phi_{ave}$	
	Invariant	Other	Invariant	Other
46	Leu 12	Ile 1, Arg 1	2.40	2.95, 0 75
48	Ile 12	Met 2	2.95	1.30

Invariant with a subgroup exception.

Position	Amino Acids		Hydrophobicity, $H\phi_{ave}$	
	Invariant	Other	Invariant	Other
8	Pro 58	Ala 6 (Human λ III)	2.60	0.75
12	Ser 56	Pro 4 (Human κ II)	0.00	2.60
		Ala 1 (Mouse κ II)		0.75
37	Glu 12, Glx 2	Pro 2 (Human κ II)	0.00	2.60

* Numbers next to the residue represent the number of samples in which the residue occurred.

It is of interest that 15 of the invariant residues have values of 0.00 or 0.45 and that 10 have values of 2.40–3.00, leaving only 4 residues of intermediate hydrophobicity: 2 half-cystines, 1 arginine, and 1 alanine. In most instances the other substituents reported as replacements at these positions generally had $H\phi_{ave}$ values not too different from the major substituent, but at positions 40, 41, 63, 67, 75, and 84 changes of over one unit were seen. (Position 41 may not be significant since, as discussed earlier, the exception had glycine in position 39.) The two borderline invariant residues both showed substantial $H\phi_{ave}$ differences and the three subgroup specific residues also varied substantially in $H\phi_{ave}$.

Of the 35 invariant residues in the constant region, 11 have values of 0.00 or 0.45, 13 have values of 2.40 to 3.00, and 11 have values of 0.75 to 1.70 (Table V). Thus the constant region has invariant residues which appear to be relatively uniformly distributed with respect to $H\phi_{ave}$ while in the variable region they are generally either very high or zero.

The average hydrophobicity for invariant residues of the variable region is about 1.09 while that of invariant residues of the constant region is 1.39.

Welcher (51) has computed $H\phi_{ave}$ for the entire light chain from sequence data on Bence Jones proteins and obtains values ranging from 0.970 to 1.04 kcal/residue, values in the same range as those for chains of other proteins. The variable regions range from 0.930 to 1.11 while the constant region values were

0.950 for mouse κ, 0.970 for human κ, and 1.02 for human λ. Thus the over-all hydrophobicities of the two regions are no different; Welcher also reported no difference between the two regions with respect to the pattern of nonpolar positions, while our data show a substantial difference in $H\emptyset_{ave}$ of the invariant residues of the two regions. Moreover, the average hydrophobicity of the entire

TABLE V

Hydrophobicity, $H\emptyset_{ave}$, of the Invariant Residues of the Constant Region

Position	Amino Acid	Hydrophobicity $H\emptyset_{ave}$	Position	Amino Acid	Hydrophobicity $H\emptyset_{ave}$
111	Ala	0.75	149	Lys	1.50
112	Ala	0.75	151	Asp	0.00
113	Pro	2.60	168	Ser	0.00
115	Val	1.70	173	Tyr	2.85
118	Phe	2.65	176	Ser	0.00
119	Pro	2.60	177	Ser	0.00
120	Pro	2.60	179	Leu	2.40
121	Ser	0.00	181	Leu	2.40
123	Glu	0.00	189	His	0.00
125	Leu	2.40	192	Tyr	2.85
130	Ala	0.75	194	Cys	1.00
133	Val	1.70	197	Thr	0.45
134	Cys	1.00	198	His	0.00
139	Phe	2.65	203	Ser	0.00
140	Tyr	2.85	207	Lys	1.50
141	Pro	2.60	213	Glu	0.00
146	Val	1.70	214	Cys	1.00
148	Trp	3.00			

constant region is about 0.98, significantly lower than 1.39 for the hydrophobicity of its invariant residues.

Variability—In considering the nature of the variable region, it is of importance to ascertain whether the variability is uniformly distributed or is confined to small segments of the variable regions. Thus, a quantity is defined for each amino acid position in the sequence

$$\text{Variability} = \frac{\text{Number of different amino acids at a given position}}{\text{Frequency of the most common amino acid at that position}} \quad [1]$$

in which the denominator is the number of times the most common amino acid occurs divided by the total number of proteins examined. Thus at position 7 (Table I) 63 proteins were studied, serine occurred 41 times and 4 different

amino acids, Pro, Thr, Ser, and Asp, have been reported. The frequency of the
most common is $41/63 = 0.65$ and the variability is then $4/0.65 = 6.15$. When
there was uncertainty as to the number of amino acids, as in instances in which
Glx or Asx has been reported, the extreme values of variability have been com-
puted. For this equation an absolutely invariant residue would have a value of
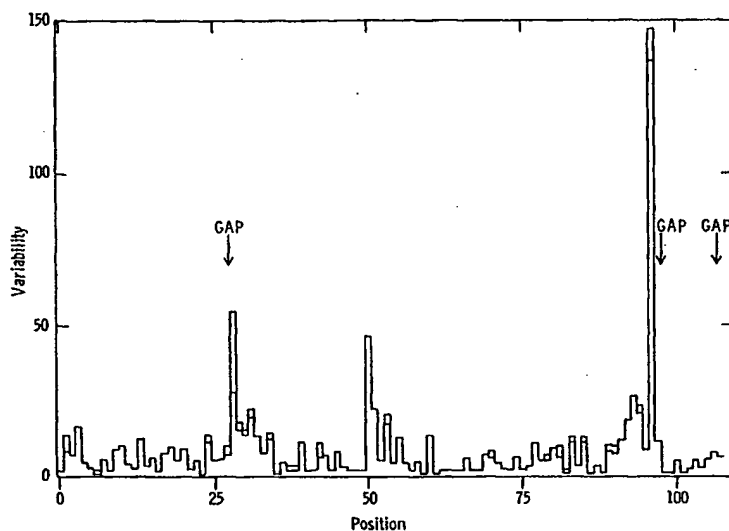1 while the theoretical upper limit for 20 amino acids randomly occurring would
be 400.

Fig. 2. Variability at different amino acid positions for the variable region of the light
chains. GAP indicates positions at which insertions have been found.

Plotting variability against position for the 107 residues of the variable
region (Fig. 2) shows three main peaks in the regions of residues 28, 50, and 96;
two of these, 28 and 96, are the highly variable regions (7, 45, 47, 49) in which
insertions occur, while position 50 has not been associated with an insertion.
Franěk (49) has previously noted the high variability around position 50. The
stretches of amino acid residues showing this high variability are 24–34, 50–56,
and 89–97. The first and third regions begin after an invariant Cys and are
followed by an invariant Trp and Phe respectively, at positions 35 and 98. The
second region begins after an almost invariant position 49 (Tyr 13/14, Phe
1/14) and is followed by an invariant position 57 (Gly 15/16, Thr 1/16) (Table
I).

The over-all sequence data were also examined to ascertain whether amino
acid substitutions at each position were reflected in changes in hydrophobicity

confined to certain stretches of the variable region. The findings which are not plotted showed that the same three regions associated with high variability were those with the greatest variation in H$\emptyset_{ave}$.

Since a portion of the variability at many positions is group- or subgroup-specific and is therefore generally not complementarity determining, variability as defined in equation [1] was computed for the individual subgroups for which sufficient sequence data were available. A plot for κI showed high variability in the stretches 24–34 and 92–96 and at residues 53 and 56. κIII shows high variability at residue 96. Data are generally insufficient even for these two subgroups and the data for other subgroups do not permit such an analysis; the λI subgroup showed unusually high variability at position 18.

Classification of Variability at Individual Positions—The sequence data available at each position (Table I) were examined in an attempt to classify the position with respect to its role in the over-all structure. Invariant residues already considered are not included. The following categories were set up. (*a*) Invariant except for one subgroup. (*b*) κ- vs. λ-specific. (*c*) κ- vs. λ-specific with some subgroup variations. (*d*) Variation in κ- and λ-subgroups. (*e*) Variation in κ-subgroups with λ relatively constant. (*f*) Variation in λ-subgroups with κ relatively constant. (*g*) Unaccountable variability. (*h*) Insufficient data. (*i*) Possible species specificity. It has not been possible to set up absolute criteria for each of these classes and some difficulties were encountered in making these assignments, especially at positions for which the sequence data were relatively sparse. Occasional substitutions compatible with point mutations have often been neglected.

Examinations of each of these categories permit some interesting inferences to be made:

(*a*) There are five positions, 8, 12, 15, 37, and 54, which are essentially invariant except that in one subgroup another amino acid occurs. Thus at position 8, 58 proteins have Pro while 6, all of which belong to the λIII subgroup, contain Ala. At the other positions, additional amino acids sometimes occur in individual proteins. These positions are considered to be part of the basic skeleton of the variable region under the control of structural genes, the subgroup differences being the result of permissible mutations.

(*b*) Six positions, 7, 33, 71, 83, 105, and 106, show predominantly κ- vs. λ-specificity. Thus at position 7, 41 human κ- and mouse κ-proteins have Ser while 20 human λ-specimens have Pro. Two exceptions occur, a κII protein with Thr instead of Ser, and a λV with Asp instead of Pro.

(*c*) At five additional positions, 1, 2, 13, 25, and 27, evidence of κ- vs. λ-specificity persists but a substitution may occur in one or more subgroups. Thus at position 1, 16 human λ-proteins of subgroups λI, λII, λIII have PCA, while λIV and λV have no amino acid, and 31 human and mouse κI and κII proteins have Asp (5 additional have Asx) while 14 human κIII specimens have Glu

(1 additional with Glx). There are several exceptions: one human κIII with Lys and one with Asp and one mouse κII with PCA.

The 11 positions in categories b and c are of importance because they show that κ- and λ-specificity is not exclusively a property of the constant region but involves the variable region as well. These findings are consistent with the data indicating that the variable and constant regions of κ-chains always go together as do the variable and constant regions of λ-chains. They are also in accord with the immunochemical studies of Ruffilli and Baglioni (60) who showed that λ-specificity extended into the variable region, and with those of Ruffilli (61) that determinants in the variable region of κ cross-react with anti λ-sera. They also suggest a continuous evolutionary association for the variable and constant regions of the κ-chains as well as for the variable and constant regions of the λ-chains. The regions showing κ- and λ-specificity appear to be distributed at the beginning and end of the variable region.

(d–f) These three categories are instances of variation in composition ascribable to subgroup variation. At 9 positions, 3, 14, 18, 19, 22, 29, 42, 51, and 79, variation ascribable to subgroups is seen in both κ and λ chains. In an additional 10 positions, 4, 9, 10, 17, 20, 21, 55, 56, 77, and 85, subgroup variation is predominantly in κ-chains with λ relatively constant, and at 15 positions, 11, 26, 39, 47, 52, 66, 69, 72, 76, 78, 80, 89, 95, 97, and 107, the subgroup variation seems to involve λ-chains with κ relatively constant. The larger number of residues placed in category f is probably a consequence of the classification of λ-chains into five subgroups while κ-chains are only divided into three subgroups. Moreover the number of samples of λ-chains is fewer than for κ-chains so that other types of variation may be masked.

(g) Unaccountable variability. At 8 positions, 28, 30–32, 93, 94, 96, and 103, the variation within each subgroup appears to be greater than can be accounted for on any known basis, especially for those subgroups for which a sufficient number of samples have been examined. It is of interest that except for residue 103 these positions are clustered in the two regions of highest variability (45) and would be brought into close proximity by the disulfide bond I_{23}–II_{88}. It is postulated that these residues may be complementarity determining and actually be involved in making contact with the antigenic determinant. These residues are also very close to the position at which insertions occur (Table I).

(h) Insufficient data. At 15 positions, 24, 34, 36, 43, 45, 53, 58, 70, 74, 81, 87, 90–92, and 104, not enough sequences are available to assign them clearly to one of the other categories. Five of the residues, 24, 34, 90–92, occur in the two regions of highest variability close to most of those with unaccountable variability. The others are fairly well spread and only one, 53, occurs in the third highly variable region. Position 18, although placed in group d shows an extraordinary variability in the λI subgroup.

(i) Species-specific residues. At positions 50 and 60 the two mouse Bence

Jones proteins examined have residues which do not correspond to any reported for the human proteins. Thus these two positions are still classifiable as species-specific. Position 96, with its extraordinary variability, although placed in category g might also technically be called species specific since the two mouse samples both contain Trp which is absent in the human samples. At all three positions the human proteins show substantial variability in the number of substitutions which can occur—9 at position 50, 5 at position 60, and 11 at position 96. It thus seems very unlikely that the apparent species specificity of these positions will persist when more mouse sequences are available.

DISCUSSION

The ability to subject the large amount of sequence data on human and mouse Bence Jones proteins and light chains to a statistical analysis has supported the earlier conclusion about the lack of species specificity in the variable regions of human and mouse Bence Jones proteins (40). The data now available indicate at most 2 or 3 such residues in the variable region as compared with 36 in the constant region. Data in other species, although limited to the first few amino terminal residues, tend to support this. The rabbit has always been considered an important exception to the view that there was little if any species specificity in the variable region, since rabbit light chains had Ala and Ile as N-terminal residues. However, the demonstration by Hood et al. (62) that the N-terminal sequence of a homogeneous rabbit antibody to the C carbohydrate of the streptococcus was Ala-Asp-Val-Val-Met-Thr-Glu-Thr-Pro-Ala-Ser-Val indicates that the rabbit has merely added an N-terminal Ala to the N-terminal Asp, the other residues then being essentially similar to those in human light chains. Thus the rabbit is not an exception to the basic evolutionary unity of light chains.

The data now available (Table II) amply support the earlier suggestions for the role of the invariant glycines of the variable region both in conferring flexibility (Fig. 1) to permit substitutions at the variable positions, and for the glycines at positions 99 and 101, together with the frequently found glycine at residue 100, in functioning as a pivot to permit optimal fitting around the antigenic determinant (18, p. 87; and 41, 43). This postulate is now strongly supported by the finding of two glycines in an analogous region of the heavy chain. The heavy chain sequences thus far available also indicate that there may be other invariant glycines in the variable region (56, 57).

The data in Table IV show an unusual distribution of invariant residues in the variable region with respect to $H\emptyset_{ave}$ in that, with a few exceptions, these residues have either a high or a low or zero value, while $H\emptyset_{ave}$ values appear to be uniformly distributed throughout the invariant residues of the constant region (Table V). The significance and structural implications of this are not clear. This finding is not evident when one examines only the over-all hydro-

phobicity, which does not differ significantly for the two halves of the chain (51).

The variability at each position shows that variability is concentrated in three regions of the molecule (49), of which two have also been noted by other workers (7, 45, 47). Examination of the basis for the variability at each position shows that variation ascribable to κ and λ or to their subgroups does not readily explain all of the variation. Seven of the eight positions at which the variability cannot be otherwise accounted for occur in two of these regions. Of the 15 positions for which sufficient sequence data were lacking to permit clear assignment, 5 occur in the same 2 regions—24–34 and 89–97 and those are the 2 regions at which additional residues are found (Table I). Much more sequence data will be required to permit unequivocal elucidation of the role of each variable position.

For the moment, if one accepts (a) the tendency of the positions of unaccountably high variability to be concentrated in the two short stretches of the chain at which insertions are found, (b) the finding that subgroup and group specificity occurs in the variable region and indeed probably over substantial portions of it, one might formulate the following working hypothesis extending the earlier concept (42, 45): The light chains of immunoglobulins except for the regions of unaccountable variability, 24–34 and 89–97, are governed by a number of structural genes, each chain being the product of two linked genes, one for the variable and one for the constant region (Hood et al. in reference 16). These structural genes are free to mutate and are limited only by the requirement that their product be capable of assuming the proper three-dimensional structure to permit an antibody site to be formed. By hypothesis the complementarity-determining residues are considered to be the result of the insertion into the DNA of the two short linear sequences, 24–34 and 89–97, which specifically determine what kind of antibody site will be formed. In the light chain the two insertions would be brought into close proximity by the disulfide bond I_{23}–II_{88}. A similar type of insertion would be made in the heavy chain, but thus far there is evidence for only one region of high variability (56, 57).

An insertion mechanism involving only short linear sequences provides a substantial simplification of the problem of providing a seemingly limitless number of complementary sites without the use of very large amounts of DNA. It would also be more likely to provide the necessary evolutionary stability and universality for the antibody-forming mechanism which cannot be adequately accounted for by the germ line hypothesis (36) of one gene for each variable region since such a system would diverge on an evolutionary time scale. The precision with which the insertion would have to be accomplished, (e.g. changes in length of one or two nucleotides would result in nonsense) in itself would tend to prevent evolutionary divergence since failures in the insertional mechanism

would completely destroy the ability to form any antibody and individuals with such a defect would probably not survive. It eliminates difficulties in the translocation hypothesis which provides for the joining of one of many V genes with a C gene (33) and yet maintains the exclusive association of V_κ with C_κ and V_λ with C_λ genes. Indeed the finding that κ- and λ-specificity extends into the variable region supports the concept of one polypeptide chain resulting from the action of closely linked V and C genes. This also is consistent with the allotypic studies on γG, γM and γA (63–65). Moreover it has the additional merit of providing a fixed location for the antibody site while other theories (32, 34, 35, 39, 66) permit it to be formed by different portions of the variable region for various antibodies. Indeed Eisen (16) considers that three different sites, each of a different specificity, might be formed by a given V_L and V_H pair. This would make antibody sites completely different from all other sets of specific receptors known. The present hypothesis considers the site as involving a small fixed region of the molecule with specificity determined by the differences in side chains of complementarity-determining residues. This insertion hypothesis readily accounts for the variations in length occurring in these regions; no other hypothesis has explicitly considered this or adequately accounted for it.

On the basis of this hypothesis which ascribes only a role in three-dimensional folding to the first 23 N-terminal amino acid residues of the light chain, all recombinational theories of antibody formation (66, 67) become uninformative since they are based exclusively on sequence data in this region and thus are probably not dealing with a region involving antibody complementarity.

The present working hypothesis of linear regions determining antibody complementarity will stand or fall when adequate numbers of sequences for human and mouse light (and heavy) chains have been worked out.

The hypothesis makes certain predictions which also can be used to test its validity. One of these, since the insertion is hypothesized to determine complementarity, is that antibody molecules of a given specificity and with a uniform site can occur in any class or subclass of immunoglobulin by the insertion of the given short linear sequences. The finding (8) that human antidextran of α-(1 $\rightarrow$ 6) specificity may occur in γA, γM, $\gamma G2$, and sometimes in $\gamma G1$ immunoglobulins and may have κ- or λ-chains is consistent with but does not provide conclusive evidence for this hypothesis, since the human antidextran still represents mixtures with heterogeneous sizes of combining sites. More important, however, may be the findings of Pincus et al. (14) that rabbit type III and type VIII antibodies to the pneumococcal polysaccharide, which gave a straight line of slope not significantly different from 1.0 in a Sips plot for binding of an octasaccharide and could thus be considered quite homogeneous with respect to their antibody-combining sites, showed many light chains and several heavy chains on acrylamide gel electrophoresis. Thus a mixture of structurally

different antibody molecules could have the same binding affinity and therefore probably have very similar or even identical combining sites. It would be especially important to determine whether such mixtures of antibodies also belonged to different classes and had antigenically different light and heavy chains.

Further studies on homogeneous antibodies of a given specificity but of different classes or subclasses would be predicted to show similar sequences in the insertional regions if their sites were identical. Conversely, antibodies of a given specificity but showing differences in the degree of cross-reactivity with related antigens would be expected to show smaller differences in their insertion regions than would antibodies of totally unrelated specificities. Such data may soon become available as sequences on myeloma proteins with antibody activity and on homogeneous antibodies are accumulated.

While such findings would provide strong support for the hypothesis, it should be borne in mind that the contour of an antibody site having a given specificity or binding affinity for a given determinant could conceivably be formed by several kinds of patterns of amino acid sequences. Under such circumstances, antibodies which were homogeneous in binding affinity but were mixtures of molecules with different sequences in the hypothesized insertional regions would be expected not to give unique sequences in these regions. Moreover, the possibility exists that binding could be influenced to some extent by different residues adjacent to a site but not in themselves complementarity determining.

The data on idiotypic specificity of myeloma globulins (68) and antibodies (69–73) are compatible with the insertion model and with the over-all concept of antibody structure proposed. Thus idiotypic determinants which are found in the variable regions would represent antigenic determinants formed by patterns of amino acid sequences involving some of the side chains of residues from the inserted regions—namely those forming the exterior portions of the site but also including some of the residues involved in three-dimensional folding and belonging to various subgroups, etc. This could give rise to a large number of determinants generally not related to specificity but influenced by or indeed partly created by the sequence of site-determining residues. In many instances in which immunodominant groups of the idiotypic determinants were from those of the inserted regions, one might expect the same idiotypic specificity to be manifested in several classes of immunoglobulins; this has been shown to be the case for γM and γG antibodies from the same rabbit (71). Thus the findings on idiotypic specificity provide further support for the uniqueness and universality of the antibody-forming mechanism.

Several models for insertion of information into DNA have been recognized. One of these is the episomal model (74, 75), and another involves two recombinations as in P1 phage transduction (76, 77). A self-perpetuating episome

containing a large number of short nucleotide sequences the incorporation of any one of which into the structural genes of the light and heavy chains to provide the information for a given antibody complementarity provides a tempting mechanism. Such incorporation could be during embryogenesis, by which each cell receives the proper nucleotide sequences to program its structural genes for a given antibody specificity. Alternatively, if a cell is multipotent, it could have a number of sequences or the entire episome and the insertion of the proper sequence could be accomplished in some unknown manner after antigenic stimulation. While the evidence shows that one cell produces one kind of antibody at a time, and that myeloma cells produce populations of molecules which bind ligands in a homogeneous manner, such cells would already be programmed. The hypothesized programming could possibly result from the cell-cell interactions for which some evidence has been advanced. Systems of this type include transfer of information from macrophages to lymphoid cells (78) and two cell interactions such as thymus-bone marrow, etc. (16, p. 431, 79–81). There is no basis at present for any more than a passing reference to these as possibilities.

Placing the information for site complementarity in an independent self-duplicating mechanism would provide the evolutionary stability needed. It would lead to the universality which the antibody-forming system has clearly manifested, because the requirements for successful insertion would be relatively more stringent so that changes in the insertion material would have a higher frequency for completely destroying the capacity to form antibodies and such individuals would probably not survive.

There are difficulties in applying the episomal model to the antibody complementarity problem. The insertion in the antibody case would have to be by a recombination mechanism involving overlapping sequences on each side as in the P1 phage transduction, rather than as in the Campbell model. However, the degree of overlap on each side of the postulated insertions is very small, e.g., an invariant Cys at the beginning and invariant Trp 35 and Phe 98 at the end. Moreover episomes have not been found to date except in bacteria, although several possibly relevant systems in eukaryotes have been noted (74). It is also difficult to see how insertions to produce a given antibody specificity could be programmed for both the light and heavy chains since their contributions to binding are generally quite different.

The insertion model does not necessarily distinguish between germ line and somatic theories, for generating the diversity necessary to provide a large number of sites. Such diversity could be obtained by the existence of a large dictionary of insertions each of which determined a given specificity, e.g. by a germ line theory, or by the recombination of the nucleotides within the insertion material of a relatively small number of different sequences to produce a large number of recombinant sequences. This latter alternative might result in the

formation of site-determining sequences which were not exactly the same for a given antibody in different species or in different individuals. Should sequence data on homogeneous antibodies to a given antigenic determinant formed in various species show that the same sequence is complementarity determining, the dictionary model would be favored over the recombinational version. Similarly, if homogeneous antibodies to a given determinant in the same individual but belonging to the various classes (γG, γM, γA) or subclasses (γG1, γG2, etc.) of immunoglobulins have the same complementarity determining residues, a dictionary model would also be favored.

The suggestions put forward are admittedly speculative as is the case at the moment with all other hypotheses. They do however present the problem of the antibody-combining site and of immunoglobulin structure in a different way, but in a way which is susceptible to verification or disproof by further data. Whether the model proposed ultimately stands or falls, the effort to assign each amino acid residue in the variable region a definite role by the statistical analysis employed should prove useful.

SUMMARY

In an attempt to account for antibody specificity and complementarity in terms of structure, human κ-, human λ-, and mouse κ-Bence Jones proteins and light chains are considered as a single population and the variable and constant regions are compared using the sequence data available. Statistical criteria are used in evaluating each position in the sequence as to whether it is essentially invariant or group-specific, subgroup-specific, species-specific, etc.

Examination of the invariant residues of the variable and constant regions confirms the existence of a large number of invariant glycines, no invariant valine, lysine, and histidine, and only one invariant leucine and alanine in the variable region, as compared with the absence of invariant glycines and presence of three each of invariant alanine, leucine, and valine and two each of invariant lysine and histidine in the constant region. The unique role of glycine in the variable region is emphasized. Hydrophobicity of the invariant residues of the two regions is also evaluated. A parameter termed variability is defined and plotted against the position for the 107 residues of the variable region. Three stretches of unusually high variability are noted at residues 24–34, 50–56, and 89–97; variations in length have been found in the first and third of these. It is hypothesized that positions 24–34 and 89–97 contain the complementarity-determining residues of the light chain—those which make contact with the antigenic determinant. The heavy chain also has been reported to have a similar region of very high variability which would also participate in forming the antibody-combining site. It is postulated that the information for site complementarity is contained in some extrachromosomal DNA such as an episome and is incorporated by insertion into the DNA of

the structural genes for the variable region of short linear sequences of nucleotides. The advantages and disadvantages of this hypothesis are discussed.

BIBLIOGRAPHY

1. Kabat, E. A. 1966. The nature of an antigenic determinant. *J. Immunol.* **97**:1.

2. Kabat, E. A. 1968. Structural Concepts in Immunology and Immunochemistry. Holt, Rinehart & Winston, Inc., New York.

3. Goodman, J. W. 1969. Immunochemical specificity: Recent conceptual advances. *Immunochemistry.* **6**:139.

4. Cohen, S., and R. R. Porter. 1964. Structure and biological activity of immunoglobulins. *Advan. Immunol.* **4**:287.

5. Kabat, E. A. 1966. Structure and heterogeneity of antibodies. *Acta haematol.* **36**:198.

6. Cohen, S., and C. Milstein. 1967. Structure and biological properties of immunoglobulins. *Advan. Immunol.* **7**:1.

7. Milstein, C., and J. R. L. Pink. 1970. Structure and evolution of immunoglobulins. *Prog. Biophys. Mol. Biol.* **21**:209.

8. Yount, W. J., M. M. Dorner, H. G. Kunkel, and E. A. Kabat. 1968. Studies on human antibodies. VI. Selective variations in subgroup composition and genetic markers. *J. Exp. Med.* **127**:633.

9. Dorner, M. M., W. J. Yount, and E. A. Kabat. 1969. Studies on human antibodies. VII. Acrylamide gel electrophoresis of purified human antibodies and myeloma proteins, their heavy and light chains. *J. Immunol.* **102**:273.

10. Fleischman, J. B., D. G. Braun, and R. M. Krause. 1968. Streptococcal group-specific antibodies: Occurrence of a restricted population following secondary immunization. *Proc. Nat. Acad. Sci. U. S. A.* **60**:134.

11. Braun, D. G., and R. M. Krause. 1968. The individual antigenic specificity of antibodies to streptococcal carbohydrates. *J. Exp. Med.* **128**:969.

12. Eichmann, K., H. Lackland, L. Hood, and R. M. Krause. 1970. Induction of rabbit antibody with molecular uniformity after immunization with group C streptococci. *J. Exp. Med.* **131**:207.

13. Pappenheimer, A. M., Jr., W. P. Reed, and R. Brown. 1968. Quantitative studies of the specificity of anti-pneumococcal polysaccharide antibodies, types III and VIII. III. Binding of a labeled oligosaccharide derived from S8 by anti-S8 antibodies. *J. Immunol.* **100**:1237.

14. Pincus, J. H., E. Haber, M. Katz, and A. M. Pappenheimer, Jr. 1968. Antibodies to pneumococcal polysaccharides: Relation between binding and electrophoretic heterogeneity. *Science (Washington)* **162**:667.

15. Brenneman, L., and S. J. Singer. 1968. The generation of antihapten antibodies with electrophoretically homogeneous L chains. *Proc. Nat. Acad. Sci. U. S. A.* **60**:258.

16. Frisch, L., editor. 1967. Antibodies. *Cold Spring Harbor Symp. Quant. Biol.* **32**:1–603.

17. Rossi, G., T. K. Choi, and A. Nisonoff. 1969. Crystals of fragment F_{ab}': Preparation from pepsin digests of human IgG myeloma proteins. *Nature (London)* **223**:837.

18. Killander, J., editor. 1967. Gamma Globulins. *Nobel Symp.* **3**:17–643.

19. Beaumont, J. L. 1967. Une spécificité commune aux α- et β-lipoprotéines du sérum révélée par un autoanticorps de myélome-L'antigène Pg. *C. R. Acad. Sci Ser. D.* **264**:185.

20. Eisen, H. N., E. S. Simms, and M. Potter. 1968. Mouse myeloma proteins with antihapten antibody activity. The protein produced by plasma cell tumor MOPC-315. *Biochemistry.* **7**:4196.

21. Metzger, H., and M. Potter. 1968. Affinity site labeling of a mouse myeloma protein which binds dinitrophenyl ligands. *Science (Washington).* **162**:1398.

22. Potter, M., and M. A. Leon. 1968. Three IgA myeloma immunoglobulins from BALB/c mouse:Precipitation with pneumococcal C polysaccharide. *Science (Washington).* **162**:369.

23. Schubert, D., A. Jobe, and M. Cohn. 1968. Mouse myeloma producing precipitating antibody to nucleic acid bases and/or nitrophenyl derivatives. *Nature (London).* **220**:882.

24. Ashman, R. F., and H. Metzger. 1969. A Waldenström macroglobulin which binds nitrophenyl ligands. *J. Biol. Chem.* **244**:3405.

25. Cohn. M., G. Notani, and S. A. Rice. 1969. Characterization of the antibody to the C-carbohydrate produced by a transplantable mouse plasmacytoma. *Immunochemistry.* **6**:111.

26. Capra, J. D., D. Wertheimer, W. J. Yount, and H. G. Kunkel. 1969. Monoclonal γG anti-γ globulins in hypergammaglobulinemic purpura. *Clin. Res.* **17**:351.

27. American Association of Immunologists Symposium on Experimental Approaches to Homogeneous Antibody Populations. 1970. *Fed. Proc.* **29**:55–91.

28. Potter, M. 1967. The plasma cell tumors and myeloma proteins of mice. *Methods Cancer Res.* **2**:106.

29. Edelman, G. M., and J. A. Gally. 1962. The nature of Bence Jones proteins. Chemical similarities to polypeptide chains of myeloma globulins and normal γ-globulins. *J. Exp. Med.* **116**:207.

30. Quattrocchi, R., D. Cioli, and C. Baglioni. 1969. A study of immunoglobulin structure. III. An estimate of the variability of human light chains. *J. Exp. Med.* **130**:401.

31. Lennox, E. S., and M. Cohn. 1967. Immunoglobulins. *Annu. Rev. Biochem.* **36**:365.

32. Cohn, M. 1968. The molecular biology of expectation. *In* Nucleic Acids in Immunology. O. J. Plescia, and W. Braun, editors. Springer-Verlag, New York. 671.

33. Edelman, G. M., and W. E. Gall. 1969. The antibody problem. *Annu. Rev. Biochem.* **38**:415.

34. Hood, L., and D. W. Talmage. 1969. On the mechanism of antibody diversity: Evidence for the germ line basis of antibody variability. *In* Symposium on Developmental Aspects of Antibody Formation and Structure, Prague. In press.

35. Hood, L., and D. W. Talmage. 1970. Mechanism and antibody diversity: Germ line basis for variability. *Science (Washington)* **168**:325.

36. Dreyer, W. J., and J. C. Bennett. 1965. The molecular basis of antibody formation: A paradox. *Proc. Nat. Acad. Sci. U. S. A.* **54**:864.

37. Lederberg, J. 1959. Genes and antibodies. *Science (Washington).* **129**:1649.

38. Burnet, M. 1966. A possible genetic basis for specific pattern in antibody. *Nature (London).* **210**:1308.

39. Brenner, S., and C. Milstein. 1966. Origin of antibody variation. *Nature (London).* **211**:242.

40. Kabat, E. A. 1967. The paucity of species-specific amino acid residues in the variable regions of human and mouse Bence Jones proteins and its evolutionary and genetic implications. *Proc. Nat. Acad. Sci. U. S. A.* **57**:1345.

41. Kabat, E. A. 1967. A comparison of invariant residues in the variable and constant regions of human K, human L and mouse K Bence Jones proteins. *Proc. Nat. Acad. Sci. U. S. A.* **58**:229.

42. Kabat, E. A. 1968. Unique features of the variable regions of Bence Jones proteins and their possible relation to antibody complementarity. *Proc. Nat. Acad. Sci. U. S. A.***59**:613.

43. Kabat, E. A. 1969. Antibody complementarity and light chain structure. *In* Symposium on the Evolution of Immune Response. Oct. 20–22, 1969. U.S. Argonne National Laboratories, Lemont, Ill.

44. Fitch, W. M., and E. Margoliash. 1967. Construction of phylogenetic trees. *Science (Washington).* **155**:279.

45. Kabat, E. A. 1970. Heterogeneity and structure of antibody-combining sites. *Ann. N. Y. Acad. Sci.* **169**:43.

46. Phillips, D. C. 1967. The hen egg-white lysozyme molecule. *Proc. Nat. Acad. Sci. U. S. A.* **57**:484.

47. Milstein, C. 1967. Linked groups of residues in immunoglobulin K chains. *Nature (London).* **216**:330.

48. Niall, H., and P. Edman. 1967. Two structurally distinct classes of kappa-chains in human immunoglobulins. *Nature (London).* **216**:262.

49. Franěk, F. 1969. The character of variable sequences in immunoglobulins and its evolutionary origin. *In* Symposium on Developmental Aspects of Antibody Formation and Structure, Prague. In press.

50. Jukes, T. H. 1969. Evolutionary pattern of specificity regions in light chains of immunoglobulins. *Biochem. Genet.* **3**:109.

51. Welscher, H. D. 1969. Correlations between amino acid sequence and conformation of immunoglobulin light chains. I. Hydrophobicity and fraction charge. *Int. J. Protein Res.* **1**:235. II. Sequence comparison and the pattern of nonpolar residues. *Int. J. Protein Res.* **1**:267.

52. Blake, C. C. F., A. G. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma. 1967. On the conformation of the hen egg-white lysozyme molecule. *Proc. Roy. Soc. Ser. B. Biol. Sci.* **167**:365.

53. Birktoft, J. J., B. W. Matthews, and D. M. Blow. 1969. Atomic coordinates of tosyl-α-chymotrypsin. *Biochem. Biophys. Res. Commun.* **36**:131.

54. Edsall, J. T., P. L. Flory, J. C. Kendrew, A. M. Liquori, G. Némethy, G. N. Ramachandran, and H. A. Scheraga. 1966. A proposal of standard conventions and nomenclature for the description of polypeptide conformations. *Biopolymers* **4**:121.

55. Ramachandran, G. N., and V. Sasisekharan. 1968. Conformation of polypeptides and proteins. *Advan. Protein Chem.* **23**:283.

56. Press, E. M., and N. M. Hogg. 1969. Comparative study of two immunoglobulin G Fd-fragments. *Nature (London)*. **223**:807.

57. Cunningham, B. A., M. N. Pflumm, U. Rutishauser, and G. M. Edelman. 1969. Subgroups of amino acid sequences in the variable regions of immunoglobulin heavy chains. *Proc. Nat. Acad. Sci. U. S. A.* **64**:997.

58. Tanford, C. 1962. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Amer. Chem. Soc.* **84**:4240.

59. Bigelow, C. C. 1967. On the average hydrophobicity of proteins and the relation between it and protein structure. *J. Theor. Biol.* **16**:187.

60. Ruffilli, A., and C. Baglioni. 1967. Subgroups of L type Bence Jones proteins. *J. Immunol.* **98**:874.

61. Ruffilli, A. 1968. The antigenic determinants of the variable half of a Bence Jones protein of type L. *J. Immunol.* **100**:201.

62. Hood, L., H. Lackland, K. Eichmann, J. Ohms, and R. M. Krause. 1970. Amino acid sequences from rabbit antibody light chains and gene evolution. In preparation.

63. Todd, C. W. 1963. Allotypy in rabbit 19S protein. *Biochem. Biophys. Res. Commun.* **11**:170.

64. Todd, C. W. 1966. Discussion. *J. Cell. Physiol.* **67**(Suppl. *1*):95.

65. Koshland, M. E., J. J. Davis, and N. J. Fujita. 1969. Evidence for multiple gene control of a single polypeptide chain: The heavy chain of rabbit immunoglobulin. *Proc. Nat. Acad. Sci. U. S. A.* **63**:1274.

66. Smithies, O. 1963. Gamma-globulin variability: A genetic hypothesis. *Nature (London)*. **199**:1231.

67. Edelman, G. M., and J. A. Gally. 1967. Somatic recombination of duplicated genes: An hypothesis on the origin of antibody diversity. *Proc. Nat. Acad. Sci. U. S. A.* **57**:353.

68. Slater, R. J., S. M. Ward, and H. G. Kunkel. 1955. Immunological relationships among the myeloma proteins. *J. Exp. Med.* **101**:85.

69. Kunkel, H. G., M. Mannik, and R. C. Williams. 1963. Individual antigenic specificities of isolated antibodies. *Science (Washington)* **140**:1218.

70. Oudin, J., and M. Michel. 1963. Une nouvelle forme d'allotypie des globulines γ de sérum de lapin apparemment liée à la jonction et à la specificité anticorps. *C. R. Acad. Sci.* **257**:805.

71. Oudin, J., and M. Michel. 1969. Idiotypy of rabbit antibodies. I. Comparison of idiotype of antibodies against *Salmonella typhi* with that of antibodies against other bacteria in the same rabbit, or of antibodies against *Salmonella typhi* in various rabbits. *J. Exp. Med.* **130**:595. II. Comparison of idiotypy of various kinds of antibodies formed in the same rabbit against *Salmonella typhi. J. Exp. Med.* **130**:619.

72. Daugharty, H., J. E. Hooper, A. B. MacDonald, and A. Nisonoff. 1969. Quantitative investigations of idiotypic antibodies. I. Analysis of precipitating antibody populations. *J. Exp. Med.* **130**:1047.

73. Hurez, D., G. Meshaka, C. Mihaesco, and M. Seligmann. 1968. The inhibition by normal γG-globulins of antibodies specific for individual γG myeloma proteins. *J. Immunol.* **100**:69.

74. Campbell, A. M. 1962. Episomes. *Advan. Genet.* **11**:101.
75. Campbell, A. M. 1969. Episomes. Harper & Row, Publishers, New York.
76. Lennox, E. S. 1955. Transduction of linked genetic characters of the host by bacteriophage P1. *Virology.* **1**:190.
77. Ozeki, H., and H. Ikeda. 1968. Transduction mechanisms. *Annu. Rev. Genet.* **2**:245.
78. Adler, F. L., M. Fishman, and S. Dray. 1966. Antibody formation initiated *in vitro.* III. Antibody formation and allotypic specificity directed by ribonucleic acid from peritoneal exudate cells. *J. Immunol.* **97**:554.
79. Miller, J. F. A. P., and G. F. Mitchell. 1968. Cell to cell interaction in the immune response. I. Hemolysin-forming cells in neonatally thymectomized mice reconstructed with thymus or thoracic duct lymphocytes. *J. Exp. Med.* **128**:801.
80. Mitchell, G. F., and J. F. A. P. Miller. 1968. Cell to cell interaction in the immune response. II. The source of hemolysin-forming cells in irradiated mice given bone marrow and thymus or thoracic duct lymphocytes. *J. Exp. Med.* **128**:821.
81. Claman, H. N., E. A. Chaperon, and R. F. Triplett. 1966. Immunocompetence of transferred thymus-marrow cell combinations. *J. Immunol.* **97**:828.